Lie theory and the wave equation in space–time. 4. The Klein–Gordon equation and the Poincaré group

E. G. Kalnins

Mathematics Department, University of Waikato, Hamilton, New Zealand

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 6 April 1977)

A detailed classification is made of all orthogonal coordinate systems for which the Klein–Gordon equation in space–time, $\psi_{tt}-\Delta_3\psi=\lambda\psi$, admits a separation of variables. We show that the Klein–Gordon equation is separable in 261 orthogonal coordinate systems. In each case the coordinate systems presented are characterized in terms of three symmetric second order commuting operators in the enveloping algebra of the Poincaré group. This paper also consitutes an important step in the study of separation of variables for the wave equation in space–time $\psi_{tt}-\Delta_3\psi=0$, and its relation to the underlying conformal symmetry group O(4,2) of this equation.

INTRODUCTION

In this paper we continue[1–3] an investigation of the connection between separation of variables for the wave equation in space—time

$$\psi_{tt}-\Delta_3\psi=0, \tag{0.1}$$

and the O(4, 2) symmetry group of this equation. Here we study all the orthogonal coordinate systems for which the Klein—Gordon equation

$$\Box\psi=\psi_{tt}-\Delta_3\psi=\lambda\psi, \quad \lambda\neq 0 \tag{0.2}$$

admits a separation of variables. [By simply setting $\lambda=0$ in our results we will obtain orthogonal separable systems for (0.1).] The method used to compute all such coordinate systems is an adaptation of that used by Eisenhart[4] in the case of the Helmholtz equation in three-dimensional Euclidean space. The work of Eisenhart enables us to classify all distinct orthogonal differential forms

$$ds^2=\sum_{i=1}^{4} H_i^2\,dx_i^2, \tag{0.3}$$

and hence coordinate systems for which (0.2) admits a separation of variables. In (0.3) the H_i^2 are real functions of the new variables x_i such that sign $H_i^2=+$ for $i=1,2,3$, and sign $H_4^2=-$. The coordinates x_i are related to the standard space—time coordinates x, y, z, t by the real functions $G_i(x_1,x_2,x_3,x_4)$, where $t=G_1$, $x=G_2$, $y=G_3$, and $z=G_4$. In terms of the standard coordinates, the differential form (0.3) becomes

$$ds^2=dx^2+dy^2+dz^2-dt^2. \tag{0.4}$$

With each such differential form we give the associated space—time coordinate functions G_i and the expression for the Klein—Gordon equation in these coordinates. We also write out the separation equations, identifying their solutions as much as possible, and we compute the three commuting operators L_i $(i=1,2,3)$ whose eigenvalues are the separation constants. Each of these three operators is written as a symmetric second order operator in the enveloping algebra of the Poincaré symmetry group E(3, 1) of the Klein—Gordon equation (0.2).

When $\lambda=-m^2$, $m>0$, Eq. (0.2) becomes

$$(\Box+m^2)\,\psi(X)=0, \quad X=(t,x,y,z); \tag{0.5}$$

the relativistic equation describing a free neutral scalar particle with mass m. In the standard field-theoretic treatments of (0.5),[5] one expresses a positive-energy solution ψ in terms of its Fourier transform

$$\psi(X)=\frac{1}{(2\pi)^{3/2}}\times\iiint\frac{\exp[-i(tk_0-xk_1-yk_2-zk_3)]}{\sqrt{2}}\,f(\mathbf{k})\,dm(\mathbf{k}), \tag{0.6}$$

where the integration surface is the hyperboloid $k_0^2-k_1^2-k_2^2-k_3^2=m^2$, $k_0>0$. The Lebesgue measurable functions $f(\mathbf{k})$, such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|f(\mathbf{k})|^2\,dm(\mathbf{k})<\infty, \quad dm(\mathbf{k})=dk_1\,dk_2\,dk_3/k_0, \tag{0.7}$$

form a Hilbert space H_m with inner product

$$\langle f_1,f_2\rangle=\iiint_{-\infty}^{\infty} f_1(k)\overline{f_2}(k)\,dm(k), \quad f_1,f_2\in H_m. \tag{0.8}$$

The mapping (0.6) then induces a Hilbert space structure on the solution space of (0.5) given by

$$(\psi_1,\psi_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\Psi_1(x)\,\partial_t\bar{\psi}_2(x)-(\partial_t\Psi_1(x))\,\bar{\psi}_2(x)]\,dx\,dy\,dz \tag{0.9}$$

(independent of t), where ψ_j is related to $f_j\in H_m$ by (0.6). The natural action of the connected Poincaré group E(3, 1) on the functions ψ induces an action on the transform space H_m which is well known to be unitary and irreducible.[5]

In studies of this physical system it is obviously of interest to construct various orthonormal bases for H_m, particularly bases which correspond to separable solutions of (0.5). However, with few exceptions, only the plane wave basis (corresponding to separation in

 0022-2488/78/1906-1233$1.00

Cartesian coordinates) is employed in the published literature. Here we show explicitly that every orthogonal separable coordinate system for (0.5) has the property that the associated separated solutions are characterized as simultaneous eigenfunctions of a commuting triplet of second-order symmetry operators from the enveloping algebra of E(3, 1). The corresponding operators acting on the domain of C^∞ functions with compact support in H_m are obviously symmetric. These operators can then be extended to a commuting triplet of self-adjoint operators on H_m. (However, in some cases the deficiency indices are equal but nonzero, so that the extension is not unique. Furthermore, in a few cases the deficiency indices of some operators are unequal. This difficulty can be removed by extending the Hilbert space to include the negative energy solutions.) The spectral theorem for commuting sets of self-adjoint operators thus implies the existence of a basis for H_m which is a (generalized) eigenbasis of the commuting operators. Mapping the eigenbasis to the solution space of (0.5) via (0.6), we see that the basis eigenfunctions are separable solutions of (0.5). The spectral resolutions of the defining self-adjoint operators as computed in H_m can then be used to derive expansion theorems and special function identities for solutions of (0.5). Our characterization of orthogonal separable systems in terms of commuting second-order operators in the enveloping algebra which act within a unitary irreducible representation of E(3, 1) is an essential part of this program.

The paper is arranged as follows. In Sec. 1 we present the necessary details concerning the generators of the Poincaré group. In addition we give a preliminary discussion concerning the arrangement and computation of the coordinate systems. In Sec. 2 we extend the work of Eisenhart to consider orthogonal differential forms in four variables and then compute all the inequivalent classes of differential forms. In Sec. 3 we give the coordinate systems, separation equations, and operators defining the separation constants.

I. SOME PROPERTIES OF THE POINCARE GROUP E(3,1)

Here we briefly present those properties of the Poincaré group E(3, 1) that are relevant to this article. For more details concerning this group the reader is referred to paper 3 of this series and Refs. 6, 7, and 8. The Poincaré group consists of all proper real linear transformations which preserve the differential form (0.4). The group is the semidirect product of the group of translations T_4 in the space and time coordinate and the group of proper real Lorentz transformations SO(3, 1), i.e.

$$E(3,1) = T_4 \times SO(3,1).$$

The Lie algebra is ten-dimensional with basis elements:

1. Translations

$$P_0 = \partial_t;\quad P_1 = \partial_x,\quad P_2 = \partial_y,\quad P_3 = \partial_z;$$

2. Pure Lorentz transformations

$$N_1 = t\partial_x + x\partial_t,\quad N_2 = t\partial_y + y\partial_t,\quad N_3 = t\partial_z + z\partial_t;$$

3. Rotations

$$M_1 = y\partial_z - z\partial_y,\quad M_2 = x\partial_z - z\partial_x,\quad M_3 = x\partial_y - y\partial_x.$$

These generators satisfy the commutation relations

$$[M_i, M_j] = \epsilon_{ijk} M_k,\quad [M_i N_j] = \epsilon_{ijk} N_k,\quad [N_i, N_j] = -\epsilon_{ijk} M_k,$$
$$[P_i, N_j] = \delta_{ij} P_0,\quad [P_i, M_j] = \epsilon_{ijk} P_k,$$

where $i, j, k = 1, 2, 3$;

$$[P_0, N_j] = P_j,\quad [P_0, M_j] = 0,$$

for $j = 1, 2, 3$, and

$$[P_i, P_j] = 0,$$

for all i, j.

On the Hilbert space H_m [Eqs. (0.7) and (0.8)] the Lie algebra generators are

$$P_0 = -ik_0,\quad P_j = ik_j,\quad N_j = k_0\partial_{k_j},\quad i = 1, 2, 3,$$
$$M_1 = k_2\partial_{k_3} - k_3\partial_{k_2},\quad M_2 = k_1\partial_{k_3} - k_3\partial_{k_1},$$
$$M_3 = k_1\partial_{k_2} - k_2\partial_{k_1}.$$

In addition to the real Poincaré group E(3, 1) we will also consider its complexification E(4, $\mathbb{C}$). This is the group of proper complex transformations which preserve the differential form

$$ds^2 = dz_1^2 + dz_2^2 + dz_3^2 + dz_4^2,$$

where $z_i \in \mathbb{C}$, $i = 1, 2, 3, 4$. The group E(4, $\mathbb{C}$) is the semidirect product of the translation group T_4 and SO(4, $\mathbb{C}$), i.e.,

$$E(4,\mathbb{C}) = T_4 \times SO(4,\mathbb{C}).$$

The Lie algebra is ten-dimensional with basis elements:

1. Translations $P_i = \partial_{z_i}$, $i = 1, 2, 3, 4$,

2. Rotations $I_{ij} = z_i\partial_{z_j} - z_j\partial_{z_i}$,

with $i, j = 1, 2, 3, 4$ and $i \neq j$.

These basis elements satisfy

$$[I_{kl}, I_{st}] = \delta_{ls} I_{kt} - \delta_{ks} I_{lt} - \delta_{lt} I_{ks} + \delta_{kt} I_{ls},$$
$$[P_i, P_j] = 0,$$
$$[P_i, I_{kl}] = \delta_{ik} P_l - \delta_{il} P_k.$$

II. ORTHOGONAL SEPARABLE DIFFERENTIAL FORMS FOR THE KLEIN-GORDON EQUATION AND ITS COMPLEXIFICATION

In this section we classify the possible orthogonal differential forms which enable (0.2) or its complexification

$$\sum_{i=1}^{4} \partial_{z_i z_i}\psi = \lambda\psi \tag{2.1}$$

to be solved by separation of variables. By this we mean a classification of all choices of new variables x_1, x_2, x_3, x_4, such that $t = G_1$, $x = G_2$, $y = G_3$, and $z = G_4$.

In the case of the Klein–Gordon equation, the real functions G_i ($i = 1, 2, 3, 4$) are real differentiable functions of the real variables x_i ($i = 1, 2, 3, 4$). In order

that the new coordinates x_i be orthogonal we have the additional requirement that

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2 = \sum_{i=1}^{4} H_i^2\, dx_i^2, \tag{2.2}$$

where sign $H_i^2 = +$ for $i = 1, 2, 3$ and sign $H_4^2 = -$.

In the case of the complexified Klein—Gordon equation, the functions G_i $(i = 1, 2, 3, 4)$ are analytic functions of the complex variables x_i. The requirement of orthogonality is the same as in the real case but with no restrictions on the signs of the metric coefficients. The coordinate systems fall into five broad classes, whose general features we now summarize. Details of the derivations are given in Ref. 9.

A. Coordinate systems of class I

These correspond to coordinate systems giving the differential form

$$ds^2 = \frac{(x_1 - x_2)}{4}\left[\frac{dx_1^2}{x_1^2} - \frac{dx_2^2}{x_2^2}\right] + \epsilon x_1 x_2 (dx^2 + dy^2), \quad \epsilon = \pm, \tag{2.3}$$

where x, y can be replaced by one of the four possible coordinate systems in the Euclidean plane (in the case of the real Klein—Gordon equation). In the case of the complexified equation, x and y can be replaced by one of the various possible coordinate systems for all the complex Euclidean plane.[10] The separable solutions of (0.2) for coordinate systems of this type assume the typical form

$$\Psi = e^{-(a+b)}\, De_\nu(a + \tfrac{1}{2}x, \sqrt{-\lambda l_1}) \times De_\nu(b + \tfrac{1}{2}x, \sqrt{-\lambda l_1})\, E_3(x_3)\, E_4(x_4), \tag{2.4}$$

where $e^a = \sqrt{x_1}$, $e^b = \sqrt{x_2}$, $\tanh x = (\lambda + l_1)/(\lambda - l_1)$. Here x_3 and x_4 correspond to the appropriate choice of coordinates in the Euclidean plane and $\phi(x, y) = E_3(x_3)\, E_4(x_4)$ is a solution of

$$[(N_2 + M_3)^2 + (N_3 - M_2)^2]\,\phi(x, y) = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = l_1 \phi \tag{2.5}$$

and $x = x(x_3, x_4)$, $y = y(x_3, x_4)$. Furthermore, $D = S, C$ and Se_ν, Ce_ν are Mathieu functions.[11]

B. Coordinate systems of class II

These correspond to systems giving the differential form

$$ds^2 = \frac{(x_1 - x_2)}{4}\left[\frac{dx_1^2}{x_1} - \frac{dx_2^2}{x_2}\right] + x_1 x_2\, d\omega^2, \tag{2.6}$$

where $d\omega^2$ is one of the differential forms associated with the two-dimensional sphere or the two-dimensional single or double sheeted hyperboloids. The separable solutions of (0.2) for systems of this type appear as

$$\Psi = (x_1 x_2)^{-1/4} M_{\pm i(\lambda l_3)^{1/2}/2,\, 1/2(j+1/2)}(\pm i x_1/2\sqrt{\lambda}) \times M_{\pm i(\lambda l_3)^{1/2}/2,\, 1/2(j+1/2)}(\pm i x_2/2\sqrt{\lambda})\, E_3(x_3)\, E_4(x_4), \tag{2.7}$$

where $E_3(x_3)\, E_4(x_4) = \phi$ is a solution of

$$(N_1^2 + N_2^2 - M_3^2)\,\phi = j(j+1)\,\phi \tag{2.8}$$

and the coordinates x_3, x_4 are one of the nine possible types for which this equation admits a separation of variables. Here $M_{\alpha,\mu}$ is a Whittaker function.[12]

C. Coordinate systems of class III

These correspond to systems giving the differential form

$$ds^2 = \frac{(x_1 - x_2)}{4}\left[\frac{dx_1^2}{x_1(x_1 - 1)} - \frac{dx_2^2}{x_2(x_2 - 1)}\right] + x_1 x_2\, d\omega^2, \tag{2.9}$$

with $d\omega^2$ as in Class II. The separable solutions of (0.2) for systems of this type assume the typical form

$$\Psi = (x_1 x_2)^{1/4} Ps_\nu^{j+1/2}(\sqrt{1 - x_1}, \lambda)\, Ps_\nu^{j+1/2}(\sqrt{1 - x_2}, \lambda) \times E_3(x_3)\, E_4(x_4), \tag{2.10}$$

where $E_3(x_3)\, E_4(x_4)$ is as in Class II and Ps_ν^μ is a spheroidal function.[11]

D. Coordinate systems of class IV

These correspond to systems giving a differential form

$$ds^3 = \frac{(x_1 - x_2)}{4}\left[\frac{dx_1^2}{x_1(x_1 - 1)} - \frac{dx_2^2}{x_2(x_2 - 1)}\right] + x_1 x_2\, dx_3^2 + (x_1 - 1)(x_2 - 1)\, dx_4^2. \tag{2.11}$$

The separation equations are

$$4\frac{d}{dx_i}\left(x_i(x_i - 1)\frac{dE_i}{dx_i}\right) - \left(\frac{l_1}{x_i} + \frac{l_2}{x_i - 1} + \lambda x_i + l_3\right) E_i = 0 \quad (i = 1, 2) \tag{2.12}$$

$$\frac{d^2 E_3}{dx_3^2} = l_1 E_3, \quad \frac{d^2 E_4}{dx_4^2} = l_2 E_4.$$

E. Coordinate systems of class V

These correspond to systems with a differential form

$$ds^2 = (x_2 - \mu)(x_3 - \mu)(x_4 - \mu)\, dx_1^2 + \sum_{i=1}^{4} \frac{(x_i - x_j)(x_i - x_j)}{4f(x_i)}\, dx_i^2, \tag{2.13}$$

where $i, j, k = 2, 3, 4$ are distinct, $f(x)$ is a polynomial such that $1 \leq \deg f(x) \leq 3$, and $x = \mu$ is a root of $f(x)$. The separation equations are

$$4\left(\frac{f(x_i)}{(x_i - \mu)}\right)^{1/2} \frac{d}{dx_i}\left(\sqrt{(x_i - \mu) f(x_i)}\,\frac{dE_i}{dx_i}\right) + [-\lambda(x_i - \mu)^3 + l_1(x_i - \mu)^2 + l_2(x_i - \mu) + l_3]E_i = 0, \tag{2.14}$$

where $i = 2, 3, 4$ and

$$(\mu - \mu')(\mu - \mu'')\frac{d^2 E_1}{dx_1^2} = -l_3 E_1,$$

where μ' and μ'' are the other roots of $f(x)$ with multiplicity included and $\deg f(x) = 3$. Similar separation

equations exist in the variable x_1 when $\deg f(x)=2$ and 1.

F. Coordinate systems of class VI

These correspond to systems giving differential forms

$$ds^2=\sum_{i=1}^{4}(x_i-x_j)(x_i-x_k)(x_i-x_l)\frac{dx_i^2}{4f(x_i)}, \tag{2.15}$$

where $i,j,k,l=1,2,3,4$ are distinct and $f(x)$ is a polynomial of degree less than or equal to 4. The separation equations are

$$4\sqrt{f(x_i)}\,\frac{d}{dx_i}\left(\sqrt{f(x_i)}\,\frac{dE_i}{dx_i}\right)+[-\lambda x_i^3+l_1x_i^2+l_2x_i+l_3]E_i=0,$$

$i=1,2,3,4$.

The remaining coordinate systems correspond to group reductions of the type $E(3,1)\supset T_1\otimes E(2,1)\supset T_1\otimes O(2,1)$, $E(3,1)\supset T_1\otimes E(3)\supset T_1\otimes O(3)$, and $E(3,1)\supset O(3,1)$ in the case of the Klein–Gordon equation and $E(4,\mathbb{C})\supset T_1\otimes E(3,\mathbb{C})\supset T_1\otimes O(3,\mathbb{C})$, $E(4,\mathbb{C})\supset O(4,\mathbb{C})$ in the case of its complexification. These systems have been derived elsewhere[13] and we make no further evaluation of them.

III. ORTHOGONAL SEPARABLE COORDINATE SYSTEMS FOR THE KLEIN-GORDON EQUATION AND ITS COMPLEXIFICATION

In this section we supplement Sec. 2 by giving the coordinates in space–time corresponding to the differential forms presented there. In addition we give the three operators, L_1, L_2, and L_3 whose eigenvalues are the three separation constants l_1, l_2, and l_3. These operators are expressed as symmetric second order operators in the enveloping algebra of the Poincaré group or its complexification. Due to the large number of possible systems we group the coordinate systems corresponding to the differential forms of Sec. 2 into classes of systems with similar properties and make an explicit count of the number of distinct coordinate systems inequivalent under the Poincaré group. We also list the systems which separate for the complexified equation only (denoted by the symbol $\mathbb{C}$), bearing in mind that distinct real systems may be equivalent in the complex case.

A. Coordinate systems of class I

A suitable choice of coordinates (2.3) with $\epsilon=-$ and sign $(x_1x_2)=+$ is

$$\begin{aligned}&(1)\quad (t-x)^2=x_1x_2,\\ &(t^2-x^2)=x_1+x_2+x_1x_2(x_3^2+x_4^2),\\ &y=\sqrt{x_1x_2}\,x_3,\quad z=\sqrt{x_1x_2}\,x_4.\end{aligned} \tag{3.1}$$

In terms of these coordinates the Klein–Gordon equation assumes the form

$$\begin{aligned}\Box\Psi=&\frac{4}{(x_1-x_2)}\left[\frac{\partial}{\partial x_1}\left(x_1^2\frac{\partial\Psi}{\partial x_1}\right)-\frac{\partial}{\partial x_2}\left(x_2^2\frac{\partial\Psi}{\partial x_2}\right)\right],\\ &-\frac{1}{x_1x_2}\left(\frac{\partial^2\Psi}{\partial x_3^2}+\frac{\partial^2\Psi}{\partial x_4^2}\right)=\lambda\Psi.\end{aligned} \tag{3.2}$$

The separation equations for the solution $\Psi=E_1(x_1)E_2(x_2)E_3(x_3)E_4(x_4)$ are

$$\begin{aligned}&\frac{d^2E_3}{dx_3^2}=(l_1-l_2)E_3,\quad \frac{d^2E_4}{dx_4^2}=l_2E_4,\\ &4\frac{d}{dx_i}\left(x_i^2\frac{dE_i}{dx_i}\right)+\left(\frac{l_1}{x_i}-\lambda x_i+l_3\right)E_i=0,\end{aligned} \tag{3.3}$$

where $i=1,2$. The three operators whose eigenvalues are the separation constants are

$$\begin{aligned}&L_1=(N_2+M_3)^2+(N_3-M_2)^2,\quad L_2=(N_3-M_2)^2,\\ &L_3=(P_0+P_1)^2+M_1^2+M_2^2+M_3^2-N_1^2-N_2^2-N_3^2.\end{aligned} \tag{3.4}$$

A typical solution for the Klein–Gordon equation (0.2) is

$$\begin{aligned}\Psi=&e^{-(a+b)}De_\nu(a+\tfrac{1}{2}x,\sqrt{-\lambda l_1})\;De_\nu(b+\tfrac{1}{2}x_1\sqrt{-\lambda l_1})\\ &\times\exp[(l_1-l_2)^{1/2}x_3+(l_2)^{1/2}x_4],\end{aligned} \tag{3.5}$$

where $D=C,S$ and the variables are defined as in (2.4).

If $\epsilon=+$, then the corresponding coordinates are obtained from (3.1) via the transformations S: $(t,x,yz)\to(x,t,iy,iz)$ and $x_3\to ix_3$, $x_4\to ix_4$.

(2) If sign $(x_1x_2)=-$ and $\epsilon=+$, the appropriate choice of space–time coordinates is obtained from (3.1) via the transformation $T:(t,x,y,z)\to(it,ix,iy,iz)$. This transformation also gives the operators describing this second type of system when applied to formulas (3.4). If $\epsilon=-$, then the corresponding coordinates can be obtained from those for which $\epsilon=+$ by the transformations S and $x_3\to ix_3$, $x_4\to ix_4$.

The remaining coordinate systems in this class are obtained by regarding x_3,x_4 [as given in (3.1) for coordinate systems (1)–(2)] as Cartesian coordinates in a Euclidean plane. This is the plane whose corresponding E(2) Lie algebra has generators $\bar{P}_1=N_2+M_3$, $\bar{P}_2=N_3+M_2$, and $\bar{M}=M_1$ with commutation relations

$$[\bar{P}_1,\bar{M}]=\bar{P}_2,\quad [\bar{P}_2,\bar{M}]=\bar{P}_1,\quad [\bar{P}_1,\bar{P}_2]=0. \tag{3.6}$$

The new coordinates are then obtained by choosing polar, parabolic, and elliptic coordinates in the x_3,x_4 plane. The three possible types of coordinates resulting from each of these three choices are obtained by the same substitutions as used to find all the systems (1)–(2), i.e., we have two inequivalent pairs of coordinate systems in each case. In all cases the operator L_3 is given by its counterpart in systems (1)–(2), and the separation equations in the variables x_1,x_2 are as in (3.3). For each case we need only give the transformation $x_3\to f(x_3,x_4)$, $x_4\to g(x_3,x_4)$ specifying the change in coordinates together with the operators L_1,L_2.

The transformation to plane polar coordinates is given by the following.

$$(3)-(4)\quad x_3\to\sqrt{x_3}\cos x_4,\quad x_4\to\sqrt{x_3}\sin x_4 \tag{3.7}$$

The x_3,x_4 dependent part of the separable solution is typically

$$E_3(x_3)E_4(x_4)=C_{(-l_3)^{1/2}}(\sqrt{-l_1}x_3)\exp[\pm(l_2x_4)^{1/2}],$$

where $C_\nu(z)$ is a solution of Bessel's equation. The basis defining operators are

$$L_1=(N_2+M_3)^2+(N_3-M_2)^2,\quad L_2=M_1^2. \tag{3.8}$$

(5)–(6) The transformation to parabolic coordinates in the plane is given by

$$x_3\to\tfrac{1}{2}(x_3+x_4),\quad x_4\to\sqrt{-x_3x_4},\quad x_4<0<x_3. \tag{3.9}$$

The x_3, x_4 part of the separable solution is typically

$$E_3(x_3)E_4(x_4)=D_{[l_2-(l_1)^{1/2}]/(l_1)^{1/2}}\left[\pm\sqrt{-l_1x_3}(1+i)\right]$$
$$\times D_{[-l_2-(l_1)^{1/2}]/(l_1)^{1/2}}[\pm\sqrt{l_4x_4}(1+i)], \tag{3.10}$$

where $D_\nu(z)$ is a parabolic cylinder function.[12] The basis defining operators are

$$L_1=(N_2+M_3)^2+(N_3-M_2)^2,$$
$$L_2=(N_3-M_2)M_1+M_1(N_3-M_2). \tag{3.11}$$

(7)–(8) The transformation to elliptic coordinates in the plane is given by

$$x_3\to c\sqrt{x_3x_4},\quad x_4\to c\sqrt{(x_4-1)(1-x_3)} \tag{3.12}$$

with $0<x_3<1<x_4$.

The x_3, x_4 part of the separable solution is typically

$$E_3(x_3)E_4(x_4)=\begin{cases}Ce_\nu(\xi,h^2)ce_\nu(\eta,h^2),\\ Se_\nu(\xi,h^2)se_\nu(\eta,h^2),\end{cases} \tag{3.13}$$

where $x_4=\cosh^2\xi$, $x_3=\cos^2\eta$, $h^2=-l_1c^2/4$, and $l_2=-\lambda_\nu(h^2)$, $\nu=0,1,2,\cdots$. The functions $ce_\nu(z,h^2)$, $se_\nu(z,h^2)$ are periodic Mathieu functions.[11] The basis defining operators are

$$L_1=(N_2+M_3)^2+(N_3-M_2)^2,$$
$$L_2=M_1^2+\tfrac{1}{2}c^2[(N_2+M_3)^2-(N_3-M_2)^2]. \tag{3.14}$$

For the remaining systems of Class I we have to consider the complexified Klein–Gordon equation.

(9) [$\mathfrak{C}$] A suitable choice of coordinates is

$$(z_1-iz_2)^2=x_1x_2,$$
$$(z_1^2+z_2^2)=x_1+x_2+2x_1x_2(x_3+x_4)(x_3-x_4)^2,$$
$$z_3+iz_4=2i\sqrt{x_1x_2(x_3+x_4)}, \tag{3.15}$$
$$z_3-iz_4=i\sqrt{x_1x_2(x_3-x_4)^2}.$$

The complexified Klein–Gordon equation has the same form as (3.2) with $\Delta_2\psi=(\partial_{33}+\partial_{44})\psi$ replaced by $[1/4(x_3-x_4)](\partial_{33}-\partial_{44})\psi$. The separation equations in the variables x_3, x_4 are

$$\frac{d^2E_i}{dx_i^2}+(-4l_1x_i+l_2)E_i=0, \tag{3.16}$$

where $i=3,4$. A typical solution of this equation is

$$E_i=\left(x_i+\frac{l_2}{2\sqrt{-l_1}}\right)C_{1/3}\left[\frac{4\sqrt{-l_1}}{3}\left(x_i+\frac{l_2}{2\sqrt{-l_1}}\right)^{3/2}\right], \tag{3.17}$$

where $C_\nu(z)$ is a solution of Bessel's equation. The separation equations in the variables x_1, x_2 are as in (3.3) with l_1+l_2 replaced by l_1. The basis defining operators are

$$L_1=(I_{32}+iI_{31})^2+(I_{42}+iI_{41})^2,$$
$$L_2=\{I_{43},I_{32}+I_{42}+i(I_{31}+I_{41})\}$$
$$+[I_{32}-I_{42}+i(I_{31}-I_{41})]^2, \tag{3.18}$$
$$L_3=(P_1-iP_2)^2+I_{12}^2+I_{13}^2+I_{14}^2+I_{23}^2+I_{24}^2+I_{34}^2.$$

(10) [$\mathfrak{C}$] A suitable choice of coordinates is

$$(z_1-iz_2)^2=x_1x_2,$$
$$(z_1^2+z_2^2)=x_1+x_2+2x_1x_2(x_3+x_4),$$
$$(z_3+iz_4)=i\sqrt{x_1x_2}\left[\left(\frac{x_3}{x_4}\right)^{1/2}+\left(\frac{x_4}{x_3}\right)^{1/2}\right], \tag{3.19}$$
$$(z_3-iz_4)=-i\sqrt{x_1x_2x_3x_4}.$$

The complexified Klein–Gordon equation has the same form as (3.2) with $\Delta_2\psi=(\partial_{33}+\partial_{44})\psi$ replaced by

$$\frac{4}{(x_3-x_4)}\left[x_3\frac{\partial}{\partial x_3}\left(x_3\frac{\partial\psi}{\partial x_3}\right)-x_4\frac{\partial}{\partial x_4}\left(x_4\frac{\partial\psi}{\partial x_4}\right)\right].$$

The separation equations in the variables x_3, x_4 are

$$x_i\frac{d}{dx_i}\left(x_i\frac{dE_i}{dx_i}\right)+(\lambda-4l_1x_i+l_1)E_i=0, \tag{3.20}$$

where $i=3,4$. A typical solution of this equation is

$$E_i=C_\nu(2\sqrt{l_1x_3})C_\nu(2\sqrt{-l_1x_4}). \tag{3.21}$$

The separation equations in the variables x_1, x_2 are as in coordinate system (9). The operators L_1 and L_3 also are the same as in system (9). The remaining operator is

$$L_2=-I_{34}^2+[I_{32}+I_{14}+i(I_{31}+I_{42})]^2. \tag{3.22}$$

B. Coordinate systems of class II

(11)–(13) In analogy to our treatment of Class I we treat one of the coordinate systems in detail and give the transformations from which the remaining coordinate systems can be obtained. A suitable choice of coordinates of type (2.6) with sign $(x_1x_2)=+$ and $x_3<0$ is

(11)

$$t=\sqrt{x_1x_2(1-x_3)},\quad x=\sqrt{-x_1x_2x_3}\cos x_4,$$
$$y=\sqrt{-x_1x_2x_3}\sin x_4,\quad z=\tfrac{1}{2}(x_1+x_2). \tag{3.23}$$

The Klein–Gordon equation assumes the form

$$\Box\psi=\frac{4}{(x_1-x_2)}\left[\sqrt{x_2}\frac{\partial}{\partial x_2}\left(\sqrt{x_2}\frac{\partial\psi}{\partial x_2}\right)-\sqrt{x_1}\frac{\partial}{\partial x_1}\left(\sqrt{x_1}\frac{\partial\psi}{\partial x_1}\right)\right]$$
$$-\frac{1}{x_1x_2}\left(4\sqrt{1-x_3}\frac{\partial}{\partial x_3}\left(x_3\sqrt{1-x_3}\frac{\partial\psi}{\partial x_3}\right)-\frac{1}{x_3}\frac{\partial^2\psi}{\partial x_4^2}\right)$$
$$=\lambda\Psi. \tag{3.24}$$

The separation equations are

$$4\sqrt{1-x_3}\frac{d}{dx_3}\left(x_3\sqrt{1-x_3}\frac{dE_3}{dx_3}\right)-\frac{l_2}{x_3}E_3=l_1E_3,$$
$$\frac{d^2E_4}{dx_4^2}=l_2E_4,$$

$$4\sqrt{x_i}\,\frac{d}{dx_i}\left(\sqrt{x_i}\,\frac{dE_i}{dx_i}\right)+-\left(\frac{l_1}{x_i}+\lambda x_i-l_3\right)E_i=0, \tag{3.25}$$

where $i=1,2$. The three defining operators are

$$L_1=N_1^2+N_2^2-M_3^2,\quad L_2=M_3^2,$$
$$L_3=P_0N_3+N_3P_0-2M_2P_2+2M_1P_1. \tag{3.26}$$

A typical solution of the Klein–Gordon equation is

$$\Psi=(x_1x_2)^{-1/4}M_{\pm i(\lambda l_3)^{1/2}/2,1/2(j+1/2)}(\pm ix_1/2\sqrt{\lambda})$$
$$\times M_{\pm i(\lambda l_3)^{1/2}/2,1/2(j+1/2)}(\pm ix_2/2\sqrt{\lambda})$$
$$\times P_j^{\sqrt{r_2}}(\sqrt{1-x_3})\exp[\pm(l_2)^{1/2}x_4], \tag{3.27}$$

where $P_\mu^\nu(z)$ is a Legendre function. There is a further coordinate system of type (11) obtained by allowing the parameters x_i to vary in the ranges $\mathrm{sign}(x_1x_2)=-$, $x_3>1$.

(12) Coordinate systems of this type correspond to the ranges, sign $(x_1x_2)=+$, $0\leq x_3<1$, and can be obtained from systems of type (11) via the transformation $(t,x,y,z)\to(z,ix,iy,t)$.

(13) If sign $(x_1x_2)=+$, $x_3>1$ or sign $(x_1x_2)=-$, $x_3<0$, and if we make the two transformations $x_4\to ix_4$ and $(t,x,y,z)\to(iy,it,ix,z)$ in (3.19) and (3.22) we get the appropriate coordinates and basis operators. If sign $(x_1x_2)=-$ and $0<x_3<1$, then making the substitutions $x_4\to ix_4$ and $(t,x,y,z)\to(iy,ix,t,z)$ we get another set of coordinates.

We can extract the essential features of the remaining distinct coordinates of Class II from the three systems already described. There are two kinds of coordinates:

$$\text{[i]}\quad t=\sqrt{x_1x_2}\tau_1,\quad x=\sqrt{x_1x_2}\tau_2,$$
$$y=\sqrt{x_1x_2}\tau_3,\quad z=\tfrac{1}{2}(x_1+x_2). \tag{3.28}$$

If sign $(x_1x_2)=+$, then the vector $\boldsymbol{\tau}=(\tau_1,\tau_2,\tau_3)$ is parametrized by one of the nine orthogonal separable coordinate systems on the double sheeted hyperboloid $[\boldsymbol{\tau},\boldsymbol{\tau}]=\tau_1^2-\tau_2^2-\tau_3^2=1$. If sign $(x_1x_2)=-$, then the vector (τ_1,τ_2,τ_3) is parametrized by one of the nine classes of orthogonal separable coordinate systems on the single sheeted hyperboloid $[\boldsymbol{\tau},\boldsymbol{\tau}]=-1$.

$$\text{[ii]}\quad t=\tfrac{1}{2}(x_1+x_2),\quad x=\sqrt{x_1x_2}\tau_1,$$
$$y=\sqrt{x_1x_2}\tau_2,\quad z=\sqrt{x_1x_2}\tau_3 \tag{3.29}$$

parametrized by one of the two orthogonal separable coordinate systems on the sphere $\tau_1^2+\tau_2^2+\tau_3^2=1$.

For the remaining coordinate systems of Class II we need only give the 3-vector (τ_1,τ_2,τ_3) in terms of the coordinates x_3,x_4, appearing in the corresponding differential form of Sec. 2. In addition we give the operator L_2 specifying each of the separable bases together with a typical solution for $E_3(x_3)E_4(x_4)$. We note here that coordinate systems already given correspond to the following $\boldsymbol{\tau}$ vectors.

(11) $\boldsymbol{\tau}=(\cosh a,\sinh a\cos\phi,\sinh a\sin\phi)$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=1$,
$-\infty<a<\infty$, $0\leq\phi<2\pi$ $(x_3=-\sinh^2 a)$,
$\boldsymbol{\tau}=(\sinh a,\cosh a\cos\phi,\cosh a\sin\phi)$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=-1$,
$-\infty<a<\infty$, $0\leq\phi<2\pi$ $(x_3=\cosh^2 a)$. (3.30)

(12) $\boldsymbol{\tau}=(\cos\theta,\sin\theta\cos\phi,\sin\theta\sin\phi)$, $\tau_1^2+\tau_2^2+\tau_3^2=1$,
$0\leq\theta<\pi$, $0\leq\phi<2\pi$ $(x_3=\sin^2\theta)$. (3.31)

(13) $\boldsymbol{\tau}=(\cosh a\cosh b,\cosh a\sinh b,\sinh a)$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=1$,
$-\infty<a,b<\infty$ $(x_3=\cosh^2 a)$,
$\boldsymbol{\tau}=(\sinh a\cosh b,\sinh a\sinh b,\cosh a)$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=-1$,
$(x_3=-\sinh^2 a)$,
$\boldsymbol{\tau}=(\sin\theta\sinh b,\sin\theta\cosh b,\cos\theta)$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=-1$,
$(x_3=\sin^2\theta)$. (3.32)

We now proceed to the remaining coordinate systems of Class II.

(14) The corresponding choices of the vector $\boldsymbol{\tau}$ are

(a) $\tau_1+\tau_2=\sqrt{x_3}$, $\tau_1-\tau_2=(1/\sqrt{x_3}x_4^2$,
$\tau_3=x_4\sqrt{x_3}$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=1$, $x_4,x_3>0$. (3.33)

(b) The coordinates corresponding to the single sheeted hyperboloid $[\boldsymbol{\tau},\boldsymbol{\tau}]=-1$ are obtained from (3.29) via the substitution $\tau\to i\tau$ with $x_3<0$. The operator L_2 for this coordinate system is

$$L_2=(N_2+M_3)^2, \tag{3.34}$$

and a typical solution for the x_3,x_4 dependent part of the solution of (0.2) is

$$E_3(x_3)E_4(x_4)=x_3^{-1/4}K_{j+1/2}(\sqrt{-l_2/x_3})\exp[(l_2)^{1/2}x_4], \tag{3.35}$$

where $K_\nu(z)$ is a Macdonald function.

(15)–(17) The corresponding choice of the vector $\boldsymbol{\tau}$ is

(15a) $\tau_1^2=x_3x_4/a$, $\tau_2^2=(x_3-1)(x_4-1)/(a-1)$,
$\tau_3^2=(x_3-a)(a-x_4)/a(a-1)$, $[\boldsymbol{\tau},\boldsymbol{\tau}]=1$,
$1<x_3<a<x_4$. (3.36)

(15b) The coordinates on the single sheeted hyperboloid $[\boldsymbol{\tau},\boldsymbol{\tau}]=-1$ are obtained from (3.32), via the substitution $\boldsymbol{\tau}\to i\boldsymbol{\tau}$ with $x_3<0<1<x_4<a$. The operator L_2 for this coordinate system is

$$L_2=N_1^2+aN_2^2 \tag{3.37}$$

and a typical solution for the x_3,x_4 dependent part of the solution of (0.2) is

$$L_{jl_2}(x_3)L_{jl_2}(x_4), \tag{3.38}$$

where $L_{jl}(z)$ is a solution of Lamé's equation

$$\frac{d^2L_{jl}}{dz^2}+\frac{1}{2}\left(\frac{1}{z-a}+\frac{1}{z-1}+\frac{1}{z}\right)\frac{dL_{jl}}{dz}$$
$$+\frac{(l_2-j(j+1)z)L_{jl}}{4(z-a)(z-1)z}=0. \tag{3.39}$$

(16a) This coordinate system is obtained from (15a)

via the transformation $(\tau_1, \tau_2, \tau_3) \to (i\tau_2, i\tau_1, \tau_3)$ and $x_3 < 0 < 1 < a < x_4$.

(16b) This coordinate system is related to (15b) in the same way as (16a) and $1 < x_3x_4 < a$, or $x_3x_4 > a$.

(17) Finally, the one system on the sphere is obtained from (15a) via the substitution $(\tau_1\tau_2\tau_3) \to (\tau_1, i\tau_2, i\tau_3)$ and $0 < x_3 < 1 < x_4 < a$.

(18) A suitable choice of coordinates on the double sheeted hyperboloid is:

$$\text{(18a)}\quad (\tau_1 + i\tau_2)^2 = 2(x_3 - a)(x_4 - a)/a(a - b),$$
$$\tau_3^2 = x_3x_4/ab, \quad [\tau, \tau] = 1, \tag{3.40}$$

and $x_3 < 0 < x_4$.

(18b) The coordinates on the single sheeted hyperboloid are obtained from those of (3.36) by the substitution $\tau \to i\tau$. The operator L_2 is

$$L_2 = \alpha(M_1^2 - N_2^2) + \beta(M_1N_2 + N_2M_1) \tag{3.41}$$

and a typical solution for the x_3, x_4 dependent part of the solution of (0.2) is

$$\tilde{L}_{jl_2}(x_3)\,\tilde{L}_{jl_2}(x_4), \tag{3.42}$$

where $\tilde{L}_{jl_2}(z)$ is a solution of

$$\frac{d^2\tilde{L}_{jl_2}}{dz^2} + \frac{1}{2}\left(\frac{1}{z-a} + \frac{1}{z-b} + \frac{1}{z}\right)\frac{d\tilde{L}_{jl_2}}{dz} + \frac{[l_2 - j(j+1)z]\tilde{L}_{jl_2}}{4(z-a)(z-b)z} = 0. \tag{3.43}$$

(19)—(20) A suitable choice of coordinates on the double sheeted hyperboloid is:

$$\text{(19a)}\quad \tau_1 + \tau_2 = \sqrt{-x_3x_4},$$
$$\tau_1 - \tau_2 = \sqrt{-x_3/x_4} + \sqrt{-x_4/x_3} - \sqrt{-x_3x_4},$$
$$\tau_3 = \sqrt{(1-x_3)(x_4-1)}, \quad [\tau, \tau] = 1 \tag{3.44}$$

and $x_3 < 0 < 1 < x_4$.

(19b) The corresponding coordinates on the single sheeted hyperboloid are obtained via the substitution $\tau \to i\tau$ with $x_3, x_4 < 0$, $0 < x_3, x_4 < 1$, $x_3, x_4 > 1$. The operator for this system is

$$L_2 = N_1^2 - (N_2 + M_3)^2 \tag{3.45}$$

and a typical solution for the x_3, x_4 dependent part is

$$E_3(x_3)\,E_4(x_4) = P_j^{(l_2)^{1/2}}(\sqrt{1-x_3})\,P_j^{(l_2)^{1/2}}(\sqrt{1-x_4}), \tag{3.46}$$

where $P_\mu^\nu(z)$ is a Legendre function.

(20a) This system is obtained from (19a) via the transformation $(\tau_1\tau_2\tau_3) \to (i\tau_2, i\tau_1, \tau_3)$ and $x_3 < 0 < 1 < x_4$.

(20b) The coordinates on the single sheeted hyperboloid are obtained from (20a) via the substitution $\tau \to i\tau$ with $x_3 < 0 < x_4 < 1$.

(21) A suitable choice of coordinates on the double sheeted hyperboloid is:

$$\text{(21a)}\quad \tau_1 + \tau_2 = \sqrt{x_3x_4}, \quad \tau_1 - \tau_2 = (x_3 - x_4)^2/4(-x_3x_4)^{3/2},$$
$$\tau_3 = \frac{1}{2}\left[\left(\frac{-x_3}{x_4}\right)^{1/2} - \left(\frac{-x_4}{x_3}\right)^{1/2}\right], \quad [\tau, \tau] = 1, \tag{3.47}$$

and $x_4 < 0 < x_3$.

(21b) The coordinates on the single sheeted hyperboloid are obtained via the substitution $\tau \to i\tau$ with $x_3, x_4 < 0$ or $x_3, x_4 > 0$. The operator for this system is

$$L_2 = N_1(N_2 - M_3) + (N_2 - M_3)N_1 \tag{3.48}$$

and a typical x_3, x_4 dependent part of the solution is

$$E_3(x_3)\,E_4(x_4) = (-x_3x_4)^{-1/4}C_{j+1/2}(\sqrt{l_2/x_3})\,C_{j+1/2}(\sqrt{-l_2/x_4}).$$

C. Coordinate systems of class III

These systems are similar to systems of Class II in that the various different types are specified by the various choices of separable coordinate systems on the manifolds $[\tau, \tau] = \pm 1$ and $\tau_1^2 + \tau_2^2 + \tau_3^2 = 1$, where $\tau = (\tau_1, \tau_2, \tau_3)$. We examine in detail one system, then discuss the general form of the coordinates in this class.

(22)—(25)

(22a) A suitable choice of coordinates with $x_3 < 0$, $x_1, x_2 > 1$; $0 < x_1, x_2 < 1$; or $x_1, x_2 < 0$ is

$$t = \sqrt{x_1x_2(1-x_3)}, \quad x = \sqrt{-x_1x_2x_3}\,\cos x_4,$$
$$y = \sqrt{-x_1x_2x_3}\,\sin x_4, \quad x = \sqrt{(1-x_1)(1-x_2)}. \tag{3.49}$$

The Klein—Gordon equation becomes

$$\Box\psi = \frac{4}{(x_1-x_2)}\left[\left(\frac{x_1-1}{x_1}\right)^{1/2}\frac{\partial}{\partial x_1}\left(x_1\sqrt{x_1(x_1-1)}\,\frac{\partial\psi}{\partial x_1}\right) - \left(\frac{x_2-1}{x_2}\right)^{1/2}\frac{\partial}{\partial x_2}\left(x_2\sqrt{x_2(x_2-1)}\,\frac{\partial\psi}{\partial x_2}\right)\right] - \frac{1}{x_1x_2}\left[4\sqrt{1-x_3}\,\frac{\partial}{\partial x_3}\left(x_3\sqrt{1-x_3}\,\frac{\partial\psi}{\partial x_3}\right) - \frac{1}{x_3}\frac{\partial^2\psi}{\partial x_4^2}\right] = \lambda\psi. \tag{3.50}$$

The separation equations in the variables x_3, x_4 are as in (3.25). The corresponding equations for the variables x_1, x_2 are

$$4\left(\frac{x_i-1}{x_i}\right)^{1/2}\frac{d}{dx_i}\left(x_i\sqrt{x_i(x_i-1)}\,\frac{dE_i}{dx_i}\right) + \left(\frac{l_1}{x_i} - \lambda x_i + l_3\right)E_i = 0, \tag{3.51}$$

where $i = 1, 2$. The three defining operators are

$$L_1 = N_1^2 + N_2^2 - M_3^2, \quad L_2 = M_3^2,$$
$$L_3 = P_0^2 - P_1^2 - P_2^2 + M_2^2 + M_1^2 - N_3^2. \tag{3.52}$$

A typical solution of the Klein—Gordon equation is

$$\psi = (x_1x_2)^{1/4}\,Ps_\nu^{j+1/2}(\sqrt{1-x_1}, -\lambda)\,Ps_\nu^{j+1/2}(\sqrt{1-x_2}, -\lambda) \times P_j^{i(l_2)^{1/2}}(\sqrt{1-x_3})\exp[\pm(l_2)^{1/2}x_4]. \tag{3.53}$$

There is a further system obtained by allowing the x_i to vary in the ranges $x_1 < 0 < 1 < x_2$, $1 > x_3 > 0$.

(22b) Systems of this type correspond to the ranges $x_1 < 0 < 1 < x_2$, $x_3 < 0$; and $0 < x_1 < 1 < x_2$, $x_3 > 1$. These systems are related to (22a) via the transformation T: $(t, x, y, z) \to (it, ix, iy, iz)$.

(23a) Systems of this type correspond to the ranges

$x_1, x_2 < 0$; $0 < x_1, x_2 < 1$; $x_1, x_2 > 1$; $0 < x_3 < 1$. These systems are related to (22a) via the transformation $(t, x, y, z) \to (z, ix, iy, t)$.

(23b) In this case we have the ranges $x_1 < 0 < 1 < x_2$ and $0 < x_3 < 1$. This is related to (23a) via T.

As for Class II, systems of Class III are of six different kinds:

[i] $t = \sqrt{x_1 x_2}\,\tau_1$, $x = \sqrt{x_1 x_2}\,\tau_2$, $y = \sqrt{x_1 x_2}\,\tau_3$,
$z = \sqrt{(1-x_1)(1-x_2)}$, $[\tau, \tau] = 1$,
and $x_1, x_2 < 0$; $0 < x_1, x_2 < 1$]; $x_1 x_2 > 1$.

[ii] $t = \sqrt{-x_1 x_2}\,\tau_1$, $x = \sqrt{-x_1 x_2}\,\tau_2$, $y = \sqrt{-x_1 x_2}\,\tau_3$,
$z = \sqrt{(1-x_1)(x_2-1)}$, $[\tau, \tau] = 1$,
and $x_1 < 0 < 1 < x_2$.

[iii] $t = \sqrt{-x_1 x_2}\,\tau_1$, $x = \sqrt{-x_1 x_2}\,\tau_2$, $y = \sqrt{-x_1 x_2}\,\tau_3$,
$z = \sqrt{(1-x_1)(1-x_2)}$, $[\tau, \tau] = -1$,
and $x_1 < 0 < x_2 < 1$.

[iv] $t = \sqrt{x_1 x_2}\,\tau_1$, $x = \sqrt{x_1 x_2}\,\tau_2$, $y = \sqrt{x_1 x_2}\,\tau_3$,
$z = \sqrt{(1-x_1)(x_2-1)}$, $[\tau, \tau] = -1$,
and $0 < x_1 < 1 < x_2$.

[v] $t = \sqrt{(1-x_1)(1-x_2)}$, $x = \sqrt{x_1 x_2}\,\tau_1$,
$y = \sqrt{x_1 x_2}\,\tau_2$, $z = \sqrt{x_1 x_2}\,\tau_2$, $\tau_1^2 + \tau_2^2 + \tau_3^2 = 1$,
and $x_1 x_2 < 0$; $0 < x_1, x_2 < 1$; $x_1, x_2 > 1$.

[vi] $t = \sqrt{(1-x_1)(x_2-1)}$, $x = \sqrt{-x_1 x_2}\,\tau_1$,
$y = \sqrt{-x_1 x_2}\,\tau_2$, $z = \sqrt{-x_1 x_2}\,\tau_3$, $\tau_1^2 + \tau_2^2 + \tau_3^2 = 1$,
and $x_1 < 0 < 1 < x_2$.

The remaining coordinate systems of Class III can be obtained from these kinds by replacing τ with the possible separable coordinate systems on the manifolds $[\tau, \tau] = \pm 1$ and $\tau_1^2 + \tau_2^2 + \tau_3^2 = 1$, exactly as for Class II. Systems (26)—(60) are of this type.[9]

D. Coordinate systems of class IV

(61) If $x_1 > 1 > x_2 > 0$, $x_3^2, x_4^2 > 0$, a suitable choice of coordinates is

$$t = \sqrt{(x_1-1)(1-x_2)}\,\sinh x_4, \quad x = \sqrt{(x_1-1)(1-x_2)}\,\cosh x_4,$$
$$y = \sqrt{x_1 x_2}\,\cos x_3, \quad z = \sqrt{x_1 x_2}\,\sin x_3. \tag{3.54}$$

The Klein—Gordon equation assumes the form

$$\Box\psi = \frac{1}{4(x_1-x_2)}\left[\frac{\partial}{\partial x_1}\left(x_1(x_1-1)\frac{\partial}{\partial x_1}\right) - \frac{\partial}{\partial x_2}\left(x_2(x_2-1)\frac{\partial\psi}{\partial x_2}\right)\right] + \frac{1}{x_1 x_2}\frac{\partial^2\psi}{\partial x_3^2} + \frac{1}{(x_1-1)(x_2-1)}\frac{\partial^2\psi}{\partial x_4^2} = \lambda\psi. \tag{3.55}$$

The separation equations are (2.12). The three defining operators are

$$L_1 = M_1^2, \quad L_2 = N_1^2,$$
$$L_3 = M_1^2 + M_2^2 + M_3^2 - N_1^2 - N_2^2 - N_3^2 + \tfrac{1}{2}[P_1^2 - P_0^2 - P_2^2 - P_3^2]. \tag{3.56}$$

Coordinates of this type are generalizations of spheroidal coordinates in three dimensions.

(62) $(t, x, y, z) \to (iy, z, x, it)$, $x_1 > 1 > x_2 > 0$, $x_3^2, x_4^2 < 0$.

(63) $(t, x, y, z) \to (ix, it, y, z)$,
$x_1, x_2 > 1$, $1 > x_1, x_2 > 0$, $0 > x_1, x_2$, $x_3^2, x_4^2 > 0$.

(64) $(t, x, y, z) \to (x, t, iy, iz)$, $x_1 > 1 > 0 > x_2$, $x_3^2, x_4^2 > 0$.

E. Coordinate systems of class V

(65)—(73)

(65) This first type corresponds to $f(x) = 4(x-a)(x-1)x$, $\mu = 0$, and

$$t^2 = \frac{(x_2-a)(x_3-a)(x_4-a)}{a(1-a)}, \quad x^2 = \frac{x_2 x_3 x_4}{a}\cos^2 x_1,$$
$$y^2 = \frac{x_2 x_3 x_4}{a}\sin^2 x_1, \quad z^2 = \frac{(x_2-1)(x_3-1)(x_4-1)}{(1-a)}. \tag{3.57}$$

In terms of these coordinates the Klein—Gordon equation becomes

$$\Box\psi = \frac{1}{(x_2-x_3)(x_2-x_4)x_2}\frac{\partial^2\psi}{\partial\nu_2^2} + \frac{1}{(x_3-x_2)(x_3-x_4)x_3}\frac{\partial^2\psi}{\partial\nu_3^2} + \frac{1}{(x_4-x_2)(x_4-x_3)x_4}\frac{\partial^2\psi}{\partial\nu_4^2} + \frac{a}{x_2 x_3 x_4}\frac{\partial^2\psi}{\partial x_1^2} = \lambda\psi, \tag{3.58}$$

where

$$\frac{\partial}{\partial\nu_j} = 2x_j\sqrt{(x_j-a)(x_j-1)}\,\frac{\partial}{\partial x_j}.$$

The three defining operators are

$$L_1 = -P_0^2 + (a+1)(P_1^2 + P_2^2) + aP_3^2 + M_1^2 + M_2^2 + M_3^2 - N_1^2 - N_2^2 - N_3^2,$$
$$L_2 = a(P_1^2 + P_2^2 + M_1^2 + M_2^2) - N_1^2 - N_2^2 + (a+1)M_3^2,$$
$$L_3 = -aM_3^2. \tag{3.59}$$

The coordinates x_2, x_3, and x_4 can vary in the ranges $x_2, x_3 > a > 1 > x_4 > 0$; $a > x_2, x_3 > 1 > x_4 > 0$; $1 > x_2 > 0 > x_3, x_4$; and $1 > x_2, x_3, x_4 > 0$ with $x_1^2 > 0$ in all cases. For the remaining systems we give the appropriate transformation of the space—time coordinates which relates the system in question to (65).

(66) $(t, x, y, z) \to (it, ix, iy, iz)$

(67) $(t, x, y, z) \to (z, ix, iy, t)$

(68) $(t, x, y, z) \to (iz, x, y, it)$

(69) $(t, x, y, z) \to (iy, x, it, z)$

(70) $(t, x, y, z) \to (iy, it, x, z)$

(71) $(t,x,y,z) \to (y,ix,t,iz)$

(72) $(t,x,y,z) \to (y,t,ix,iz)$

(73) $(t,x,y,z) \to (iy,t,ix,iz)$

(74)–(81)

(74) This type corresponds to the choice $\mu = 1$ and

$$t^2 = \frac{(x_2-a)(x_3-a)(x_4-1)}{a(1-a)},$$
$$x^2 = \frac{(x_2-1)(x_3-1)(x_4-1)}{(1-a)}\cos^2 x_1, \tag{3.60}$$
$$y^2 = \frac{(x_2-1)(x_3-1)(x_4-1)}{(1-a)}\sin^2 x_1, \quad z^2 = \frac{x_2x_3x_4}{a}.$$

The three operators are

$$\begin{aligned} L_1 &= P_0^2 + (a-2)(P_1^2+P_2^2) + (a-1)P_3^2, \\ &\quad + M_1^2 + M_2^2 + M_3^2 - N_1^2 - N_2^2 - N_3^2, \\ L_2 &= (a-1)(P_1^2+P_2^2-M_1^2-M_2^2) - N_1^2 - N_2^2 + (a-2)M_3^2, \\ L_3 &= (1-a)M_3^2. \end{aligned} \tag{3.61}$$

The coordinates can vary in the ranges

$$x_2, x_3 > a > x_4 > 1; \quad x_2, x_3 > a > 1 > x_4 > 0;$$
$$a > x_2, x_3 > 1 > x_4 > 0; \quad 1 > x_2 > 0 > x_3, x_4;$$

$1 > x_2, x_3, x_4 > 0$ where $x_1^2 > 0$ in all cases. The remaining systems are specified by the transformation of space–time coordinates which relates them to (74). The various possibilities are:

(75) $(t,x,y,z) \to (iz,x,y,it)$

(76) $(t,x,y,z) \to (z,ix,iy,t)$

(77) $(t,x,y,z) \to (it,ix,iy,iz)$

(78) $(t,x,y,z) \to (iy,x,it,z)$

(79) $(t,x,y,z) \to (y,ix,t,iz)$

(80) $(t,x,y,z) \to (y,t,ix,iz)$

(81) $(t,x,y,x) \to (iy,it,x,z)$

(82)–(86)

(82) This type corresponds to $f(x) = (x-a)(x-b)x$, $a = b^* = \alpha + i\beta$, $\alpha, \beta \in \mathbb{R}$, $\mu = 0$:

$$(z+it)^2 = \frac{2(x_2-a)(x_3-a)(x_4-a)}{a(a-b)},$$
$$x^2 = \frac{x_2x_3x_4}{ab}\cos^2 x_1, \quad y^2 = \frac{x_2x_3x_4}{ab}\sin^2 x_1. \tag{3.62}$$

The three basis defining operators are

$$\begin{aligned} L_1 &= 2\alpha(P_1^2+P_2^2) + \alpha(P_3^2-P_0^2) - 2\beta P_0P_3 \\ &\quad + M_1^2 + M_2^2 + M_3^2 - N_1^2 - N_2^2 - N_3^2, \\ L_2 &= \alpha(N_1^2+N_2^2-M_1^2-M_2^2) - (\alpha^2+\beta^2)(P_1^2+P_2^2) \\ &\quad + \beta(\{N_2,M_1\} - \{N_1,M_2\}), \quad L_3 = abM_3^2, \end{aligned} \tag{3.63}$$

where $\{A,B\} = AB + BA$.

The coordinates x_j $(j = 2,3,4)$ can vary in the ranges $x_2, x_3, x_4 > 0$; $x_2 > 0 > x_3, x_4$; with $x_1^2 > 0$.

(83) The systems of this type are obtained from (82) via the transformation T. The coordinates vary in the ranges

$$x_2, x_3 > 0 > x_4; \; 0 > x_2, x_3, x_4; \text{ with } x_1^2 > 0.$$

(84)–(87)

(84) This type corresponds to $f(x) = (x-1)x^2$, $\mu = 0$, and

$$\begin{aligned} &(t-x)^2 = x_2x_3x_4, \\ &(t^2-x^2) = -(x_4x_2 + x_4x_3 + x_2x_3) + x_4x_2x_3(1+x_1^2), \\ &y^2 = x_1x_2x_3x_4, \quad z^2 = (x_2-1)(x_3-1)(x_4-1). \end{aligned} \tag{3.64}$$

In terms of these coordinates the Klein–Gordon equation becomes (3.58) with $a = 1$ and $\partial/\partial\nu_j = 2x_j\sqrt{x_j(x_j-1)}(\partial/\partial x_j)$. The three operators are

$$\begin{aligned} L_1 &= 2P_0(P_0+P_1) + P_2^2 + M_1^2 + M_2^2 + M_3^2 - N_1^2 - N_2^2 - N_3^2, \\ L_2 &= (P_0+P_1)^2 - N_1^2 - 2N_2^2 + (N_3+M_2)^2 - N_2M_3 - M_3N_2, \\ L_3 &= (N_2-M_3)^2. \end{aligned} \tag{3.65}$$

The coordinates x_2, x_3, and x_4 can vary in the ranges $x_2, x_3, x_4 > 1$; $x_2 > 1 > x_3, x_4 > 0$; with $x_1^2 > 0$.

(85) $(t,x,y,z) \to (it,ix,iy,iz)$

(86) $(t,x,y,z) \to (x,t,iy,iz)$

(87) $(t,x,y,z) \to (ix,it,y,z)$

(88)–(90)

(88) This type corresponds to $f(x) = (x-1)x^2$, $\mu = 1$, and

$$\begin{aligned} &(t-z)^2 = x_2x_3x_4, \\ &(t^2-z^2) = -(x_1x_2 + x_1x_3 + x_2x_3) + x_1x_2x_3, \\ &x^2 = (x_2-1)(x_3-1)(x_4-1)\cos^2 x_1 \\ &y^2 = (x_2-1)(x_3-1)(x_4-1)\sin^2 x_1. \end{aligned} \tag{3.66}$$

In terms of these coordinates the Klein–Gordon equation becomes (3.58) with $1-a$ replaced by 1 and $\partial/\partial\nu_j = 2(x_j-1)\sqrt{x_j(x_j-1)}(\partial/\partial x_j)$.

The basis defining operators are

$$\begin{aligned} L_1 &= -2(P_1^2+P_2^2) + 2P_0(P_0+P_3) + M_1^2 + M_2^2 + M_3^2 \\ &\quad - N_1^2 - N_2^2 - N_3^2, \\ L_2 &= -(P_1^2+P_2^2) + 2(M_3^2 - N_1^2 - N_2^2) + \{N_2,M_1\} - \{N_1,M_2\}, \\ L_3 &= M_3^2. \end{aligned} \tag{3.67}$$

The coordinates x_2, x_3, x_4 vary in the ranges $x_2, x_3, x_4 > 1$; $x_2 > 1 > x_3, x_4 > 0$; $x_2 > 1 > 0 > x_3, x_4$; $x_2 > 1 > x_3 > 0 > x_4$; with $x_1^2 > 0$.

(89) $(t,x,y,z) \to (it,ix,iy,iz)$

(90) $(t,x,y,z) \to (z,ix,iy,t)$

(91)–(92)

(91) This system corresponds to $f(x)=x^3$, $\mu=0$, and

$$(t-x)^2=x_2x_3x_4,\quad 2y(t-x)=x_2x_3+x_2x_4+x_3x_4,$$
$$x^2+y^2+z^2-t^2=x_2+x_3+x_4,\quad z^2=x_1^2x_2x_3x_4. \tag{3.68}$$

In terms of these coordinates the Klein–Gordon equation assumes the form (3.58) with $a=1$ and $\partial/\partial\nu_j=2x_j^2(\partial/\partial x_j)$.

The three defining operators are

$$L_1=-2P_2(P_0+P_1)+M_1^2+M_2^2+M_3^2-N_1^2-N_2^2-N_3^2,$$
$$L_2=-(P_0+P_1)^2+\{M_3-N_2,N_1\}-\{M_2+N_3,M_1\},$$
$$L_3=(N_3+M_2)^2.$$

The coordinates x_j $(j=2,3,4)$ vary in the ranges $x_2,x_3>0>x_4$; and $x_2>0>x_3,x_4$; with $x_1^2>0$.

(92) $(t,x,y,z)\to(it,ix,iy,iz)$

(93)–(96)

(93) This system corresponds to $f(x)=x(x-1)$, $\mu=0$, and

$$t=\tfrac{1}{2}(x_2+x_3+x_4),\quad x^2=x_2x_3x_4\cos^2x_1,$$
$$y^2=x_2x_3x_4\sin^2x_1,\quad z^2=-(x_2-1)(x_3-1)(x_4-1). \tag{3.69}$$

In terms of these coordinates the Klein–Gordon equation becomes (3.58) with $a=1$ and $\partial/\partial\nu_j=2x_j\sqrt{x_j-1}(\partial/\partial x_j)$.

The three operators are

$$L_1=\{P_3,N_3\}-\{P_1,N_1\}-\{P_2,N_2\}+P_0^2+P_3^2,$$
$$L_2=\{N_1,P_1\}+\{N_2,P_2\}+P_1^2+P_2^2-M_1^2-M_2^2-M_3^2, \tag{3.70}$$
$$L_3=-M_3^2.$$

The coordinates x_2, x_3, and x_4 vary in the ranges $x_2,x_3>1>x_4>0$; $1>x_2,x_3,x_4>0$; $1>x_2>0>x_3,x_4$; with $x_1^2>0$.

(94) $(t,x,y,z)\to(y,t,ix,iz)$

(95) $(t,x,y,z)\to(z,ix,iy,t)$

(96) $(t,x,y,z)\to(y,ix,t,iz)$

(97)–(98)

(97) This system corresponds to $f(x)=x^2$, $\mu=0$, and

$$(t-x)^2=x_2x_3x_4,$$
$$t^2-x^2=x_2x_3+x_2x_4+x_3x_4+x_2x_3x_4x_1^2, \tag{3.71}$$
$$y^2=x_2x_3x_4x_1^2,\quad z=\tfrac{1}{2}(x_2+x_3+x_4).$$

The Klein–Gordon equation assumes the form (3.58) with $a=0$ and $\partial/\partial\nu_j=2x_j^{3/2}(\partial/\partial x_j)$.

The three operators are

$$L_1=-\{P_0+P_1,N_3-M_2\}-\{P_2,M_1\}+(P_0+P_1)^2,$$
$$L_2=\{P_0+P_1,N_3+M_2\}+N_1^2+N_2^2-M_3^2, \tag{3.72}$$
$$L_3=(N_2-M_3)^2.$$

(98) $(t,x,y,z)\to(ix,it,y,z)$

(99)–(100)

(99) This system corresponds to $f(x)=x$, $\mu=0$, and

$$2(t-z)=x_2x_3+x_2x_4+x_3x_4-\tfrac{1}{2}(x_2^2+x_3^2+x_4^2),$$
$$2(z-t)=x_2+x_3+x_4, \tag{3.73}$$
$$x^2=-x_2x_3x_4\cos^2x_1,\quad y^2=-x_2x_3x_4\sin^2x_1.$$

In terms of these coordinates the Klein–Gordon equation becomes (3.58) with $a=1$ and $\partial/\partial\nu_j=2\sqrt{x_j}(\partial/\partial x_j)$.

The three operators are

$$L_1=\{N_3,P_0+P_3\}-\{P_1,N_1-M_2\}-\{P_2,N_2+M_2\}$$
$$-\tfrac{1}{4}(P_0-P_3)^2,$$
$$L_2=\tfrac{1}{2}\{P_1,N_1+M_2\}+\tfrac{1}{2}\{P_2,N_2-M_1\}$$
$$+(N_1+M_2)^2+(N_2-M_1)^2,$$
$$L_3=M_3^2, \tag{3.74}$$

and the variables are such that sign $(x_2x_3x_4)=-1$ and $x_1^2>0$.

(100) $(t,x,y,z)\to(z,ix,iy,t)$.

F. Coordinate systems of class VII

These systems correspond to the various kinds of purely elliptical coordinates for which the Klein–Gordon equation is separable. The differential form is (2.15) where $f(x)$ is at most a fourth order polynomial in x.

(101)–(108)

(101) This type corresponds to $f(x)=(x-a)(x-b)(x-1)x$, $a>b>1$, and

$$t^2=-\frac{x_1x_2x_3x_4}{ab},\quad x^2=-\frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(a-1)(b-1)}$$
$$y^2=\frac{(x_1-b)(x_2-b)(x_3-b)(x_4-b)}{(a-b)(b-1)b},$$
$$z^2=-\frac{(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-1)a}. \tag{3.75}$$

The Klein–Gordon equation becomes

$$\Box\psi=\sum_{i=1}^{4}\frac{1}{(x_i-x_j)(x_i-x_k)(x_i-x_l)}\frac{\partial^2\psi}{\partial\nu_i^2}=\lambda\psi,$$

where i,j,k,l are not equal and

$$\frac{\partial}{\partial\nu_i}=2\sqrt{(x_i-a)(x_i-b)(x_i-1)x_i}\ \frac{\partial}{\partial x_i}. \tag{3.76}$$

The three operators are

$$L_1=-N_1^2-N_2^2-N_3^2+M_1^2+M_2^2+M_3^2+(b+1)P_3^2$$
$$+(a+1)P_2^2+(a+b)P_1^2-(a+b+1)P_0^2,$$
$$L_2=(a+b)N_1^2+(a+1)N_2^2+(b+1)N_3^2$$
$$-bM_2^2-aM_3^2-M_1^2-bP_3^2$$
$$-aP_2^2-abP_1^2+(a+b+ab)P_0^2,$$
$$L_3=abN_1^2+aN_2^2+bN_3^2+abP_0^2. \tag{3.77}$$

For coordinates of this type the variables x_i can lie in the ranges $x_1>a>x_2>b>x_3>1>0>x_4$.

(102) $(t,x,y,z)\to(ix,it,y,z)$

(103) $(t, x, y, z) \to (iz, x, y, it)$

(104) $(t, x, y, z) \to (iy, x, it, z)$

(105) $(t, x, y, z) \to (it, ix, iy, iz)$

(106) $(t, x, y, z) \to (z, ix, iy, t)$

(107) $(t, x, y, z) \to (x, t, iy, iz)$

(108) $(t, x, y, z) \to (y, ix, t, iz)$

(109)–(110)

(109) This type corresponds to $f(x) = (x-a)(x-b)(x-1)x$, $a = b^* = \alpha + i\beta$, and

$$(x+it)^2 = -\frac{2(x_1-b)(x_2-b)(x_3-b)(x_4-b)}{(b-a)(b-1)b},$$
$$y^2 = -\frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(a-1)(b-1)} \tag{3.78}$$
$$z^2 = \frac{x_1x_2x_3x_4}{ab}.$$

The Klein–Gordon equation becomes (3.76). The three operators are

$$\begin{aligned} L_1 &= N_1^2 + N_2^2 + N_3^2 - M_1^2 - M_2^2 - M_3^2 \\ &\quad + (\alpha+1)(P_1^2 - P_0^2) + 2\beta P_0 P_1 + 2\alpha P_2^2 - (2\alpha+1)P_3^2, \\ L_2 &= -2\alpha M_1^2 + (\alpha+1)(N_3^2 - M_2^2) - \beta\{M_2, N_3\} \\ &\quad + \alpha(N_2^2 - M_3^2) + \beta\{M_3, N_2\} + N_1^2 \\ &\quad + \alpha(P_0^2 - P_1^2) - 2\beta P_0 P_1 - (\alpha^2+\beta^2)P_2^2 \\ &\quad - (2\alpha + \alpha^2 + \beta^2)P_3^2, \\ L_3 &= -(\alpha^2+\beta^2)M_1^2 + \alpha(N_3^2 - M_2^2) - \beta\{M_2, N_3\}. \end{aligned} \tag{3.79}$$

The variables x_i lie in the ranges $x_1 > 1 > x_2 > 0 > x_3, x_4$; $x_1 > 1 > x_2, x_3, x_4 > 0$.

(110) $(t, x, y, z) \to (it, ix, iy, iz)$

(111)–(114)

(111) This type corresponds to $f(x) = (x-a)(x-1)x^2$, $a > 1$, and

$$(t+x)^2 = x_1x_2x_3x_4/a,$$
$$(t^2 - x^2) = -(x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_4 + x_2x_3x_4) + (a+1)x_1x_2x_3x_4/a^2,$$
$$y^2 = \frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(a-1)},$$
$$z^2 = -\frac{(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{a^2(a-1)} \tag{3.80}$$

The Klein–Gordon equation becomes (3.76) with

$$\frac{\partial}{\partial \nu_i} = 2x_i\sqrt{(x_i-a)(x_i-1)}\,\frac{\partial}{\partial x_i}.$$

The three operators are

$$\begin{aligned} L_1 &= -N_1^2 - N_2^2 - N_3^2 + M_1^2 + M_2^2 + M_3^2 \\ &\quad + 2a(P_0^2 - P_1^2) + (P_0 - P_1)^2 + aP_2^2, \\ L_2 &= -2M_2^2 + \{N_3, M_2\} - (N_2 + M_3)^2 \\ &\quad + a(N_2^2 - M_1^2) + (a+1)N_1^2 + (a+1)(P_0 - P_1)^2, \\ L_3 &= -a(N_2 + M_3)^2 + aN_1^2 - (N_3 - M_2)^2 + (P_0 - P_1)^2. \end{aligned} \tag{3.81}$$

The variables x_i lie in the ranges

$x_1 > a > x_2 > 1 > x_3, x_4 > 0$;

$x_1 > a > x_2 > 1 > 0 > x_3, x_4$;

$x_1 > a > x_2, x_3, x_4 > 1$; $x_1, x_2, x_3 > a > x_4 > 1$.

(112) $(t, x, y, z) \to (ix, it, y, z)$

(113) $(t, x, y, z) \to (x, t, iy, iz)$

(114) $(t, x, y, z) \to (it, ix, iy, iz)$

(115) [𝔉]

This type corresponds to $f(x) = (x-1)^2x^2$ and

$$\begin{aligned} &(iz_1 - z_2)^2 = x_1x_2x_3x_4, \\ &z_1^2 + z_2^2 = (x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4) - 2x_1x_2x_3x_4, \\ &(iz_3 + z_4)^2 = -(x_1-1)(x_2-1)(x_3-1)(x_4-1), \\ &z_3^2 + z_4^2 = 2x_1x_2x_3x_4 + (x_1 + x_2 + x_3 + x_4) \\ &\qquad - (x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4) - 2. \end{aligned} \tag{3.82}$$

The Klein–Gordon equation becomes (3.76) with $\partial/\partial\nu_j = 2x_j(x_j-1)(\partial/\partial x_j)$.

The operators are

$$\begin{aligned} L_1 &= I_{12}^2 + I_{13}^2 + I_{14}^2 + I_{23}^2 + I_{24}^2 + I_{34}^2 \\ &\quad + (iP_1 + P_2)^2 - \tfrac{1}{4}(P_3 - iP_2)^2, \\ L_2 &= -(iI_{13} + I_{23})^2 - (iI_{14} + I_{24})^2 \\ &\quad + (iI_{14} - I_{13})^2 - (I_{34} - iI_{23})^2 \\ &\quad - I_{14}^2 - I_{13}^2 - I_{23}^2 - I_{24}^2 - 2I_{12}^2 \\ &\quad + 2(iP_1 + P_2)^2, \\ L_3 &= -(I_{24} + iI_{23} + I_{31} + iI_{14})^2 + (iI_{14} + I_{24})^2 \\ &\quad + (I_{31} + iI_{23})^2 - I_{12}^2 + (iP_1 + P_2)^2. \end{aligned} \tag{3.83}$$

(116)–(117)

(116) This type corresponds to $f(x) = (x-1)x^3$ and

$$\begin{aligned} &(x-t)^2 = x_1x_2x_3x_4, \\ &y(x-t) = -(x_2x_3x_4 + x_1x_2x_3 + x_1x_3x_4 + x_1x_2x_4) \\ &\qquad + x_1x_2x_3x_4, \\ &y^2 + x^2 - t^2 = -(x_1x_2 + x_1x_3 + x_1x_2 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4) \\ &\qquad + x_1x_2x_3x_4, \\ &z^2 = -(x_1-1)(x_2-1)(x_3-1)(x_4-1). \end{aligned} \tag{3.84}$$

The Klein–Gordon equation becomes (3.76). The operators are

$$\begin{aligned} L_1 &= -N_1^2 - N_2^2 - N_3^2 + M_1^2 + M_2^2 + M_3^2 + 2P_2(P_0 + P_1), \\ L_2 &= N_2^2 + N_3^2 - M_1^2 - \{M_1, N_3 + M_2\} + (P_0 + P_1)^2, \\ L_3 &= -(P_0 + P_1)^2 + \{N_1, N_2 - M_1\} - (N_3 + M_2)^2. \end{aligned} \tag{3.85}$$

The coordinates x_i vary in the ranges $x_1 > 1 > x_2 > 0 > x_3, x_4$; $x_1 > 1 > x_2, x_3, x_4 > 0$, and $x_1, x_2, x_3 > 1 > x_4 > 0$.

(117) $(t, x, y, z) \to (it, ix, iy, iz)$

(118) $[\mathfrak{C}]$

This type corresponds to $f(x) = x^4$ and

$$
\begin{aligned}
&(z_1 + iz_2)^2 = -x_1x_2x_3x_4,\\
&2(z_1 + iz_2)(z_3 + iz_4) = x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4,\\
&(z_1 + iz_2)(z_3 - iz_4) + (z_3 + iz_4)^2 = -x_1x_2 - x_1x_3 - x_1x_4\\
&\qquad\qquad - x_2x_3 - x_2x_4 - x_3x_4,\\
&z_1^2 + z_2^2 + z_3^2 + z_4^2 = x_1 + x_2 + x_3 + x_4.
\end{aligned}
\tag{3.86}
$$

The Klein–Gordon equation assumes the from (3.76) with $\partial/\partial\nu_j = 2x_j^2(\partial/\partial x_j)$.

The three operators are

$$
\begin{aligned}
L_1 &= -(P_3 + iP_4)^2 + (P_1 + iP_2)(P_3 - iP_4)\\
&\quad + I_{12}^2 + I_{13}^2 + I_{14}^2 + I_{23}^2 + I_{24}^2 + I_{34}^2\\
L_2 &= \tfrac{1}{2}\{I_{32} + I_{14} + i(I_{13} + I_{24}), (I_{43} + I_{12})\}\\
&\quad + (I_{42} + iI_{23})^2 - (I_{13} + iI_{14})^2 + 2(P_3 + iP_4)(P_1 + iP_2),\\
L_3 &= \{I_{32} + I_{41} + i(I_{13} + I_{42}), I_{21}\} - (P_1 + iP_2)^2.
\end{aligned}
\tag{3.87}
$$

(119)–(122)

(119) This type corresponds to $f(x) = (x - a)(x - 1)x$ and $= (x - a)(x - 1)x]$:

$$
\begin{aligned}
t &= \tfrac{1}{2}(x_1 + x_2 + x_3 + x_4),\\
x^2 &= (x_1 - a)(x_2 - a)(x_3 - a)(x_4 - a)/a(a - 1),\\
y^2 &= (x_1 - 1)(x_2 - 1)(x_3 - 1)(x_4 - 1)/(1 - a),\\
z^2 &= x_1x_2x_3x_4/a.
\end{aligned}
\tag{3.88}
$$

The Klein–Gordon equation becomes (3.76) with

$$\frac{\partial}{\partial\nu_j} = 2\sqrt{(x_j - a)(x_j - 1)x_j}\,\frac{\partial}{\partial x_j}.$$

The operators are

$$
\begin{aligned}
L_1 &= -\{P_3, N_3\} - \{P_2, N_2\} - \{P_1, N_1\}\\
&\quad + (a + 1)P_0^2 + aP_1^2 + P_2^2,\\
L_2 &= (a + 1)\{P_3, N_3\} + \{P_1, N_1\}\\
&\quad + a\{P_2, N_1\} - aP_0^2 + M_1^2 + M_2^2\\
&\quad + M_3^2 - (a + 1)(P_1^2 + P_2^2) - P_3^2,\\
L_3 &= -a\{P_3, N_3\} - a^2P_3^2 + M_2^2 + a^2M_1^2.
\end{aligned}
\tag{3.89}
$$

The coordinates x_i vary in the ranges

$x_1, x_2 > a > x_3 > 1 > x_4 > 0$;

$a > x_1 > 1 > x_2, x_3, x_4 > 0$;

$a > x_1 > 1 > x_2 > 0 > x_3, x_4$.

(120) $(t, x, y, z) \to (z, ix, ix, iy, t)$

(121) $(t, x, y, z) \to (y, ix, t, iz)$

(122) $(t, x, y, z) \to (x, t, iy, iz)$

(123) This type corresponds to $f(x) = (x - a)(x - b)x$ and

$$
\begin{aligned}
&y = \tfrac{1}{2}(x_1 + x_2 + x_3 + x_4),\\
&(t + ix)^2 = 2(x_1 - a)(x_2 - a)(x_3 - a)(x_4 - a)/a(a - b),\\
&z^2 = -x_1x_2x_3x_4/ab, \quad a = b^* = \alpha + i\beta, \quad \alpha, \beta \in \mathbb{R}.
\end{aligned}
\tag{3.90}
$$

The Klein–Gordon equation becomes (3.76) with

$$\frac{\partial}{\partial\nu_j} = 2\sqrt{(x_j - a)(x_j - b)x_j}\,\frac{\partial}{\partial x_j}.$$

The three operators are

$$
\begin{aligned}
L_1 &= \{P_3, M_1\} + \{P_1, M_3\} - \{P_0, N_2\}\\
&\quad + 2\alpha P_2^2 + \alpha(P_0^2 - P_1^2) + 2\beta P_0P_1,\\
L_2 &= -2\alpha\{P_3, M_1\} + \alpha[\{P_0, N_2\} - \{P_1, M_3\}]\\
&\quad - 2\beta\{P_1, N_3\} + M_2^2 - N_1^2 - N_2^2 - (\alpha^2 + \beta^2)P_2^2\\
&\quad + P_3^2 + 2\alpha(P_1^2 - P_0^2),\\
L_3 &= (\alpha^2 + \beta^2)\{P_3, M_1\} + (\alpha^2 + \beta^2)^2P_3^2\\
&\quad + (\alpha^2 - \beta^2)(M_2^2 - N_3^2) + \alpha\beta\{M_2, N_3\}.
\end{aligned}
\tag{3.91}
$$

The coordinates vary in the ranges $x_1, x_2, x_3 > 0 > x_4$; $x_1 > 0 > x_2, x_3, x_4$.

(124)–(125)

(124) This type corresponds to $f(x) = (x - 1)x^2$ and

$$
\begin{aligned}
&(t - x)^2 = -x_1x_2x_3x_4, \quad z = \tfrac{1}{2}(x_1 + x_2 + x_3 + x_4),\\
&(t^2 - x^2) = x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4\\
&\qquad + x_2x_3x_4 - x_1x_2x_3x_4,\\
&y^2 = -(x_1 - 1)(x_2 - 1)(x_3 - 1)(x_4 - 1).
\end{aligned}
\tag{3.92}
$$

The Klein–Gordon equation becomes (3.76) with $\partial/\partial\nu_j = 2x_j\sqrt{x_j - 1}(\partial/\partial x_j)$.

The three operators are

$$
\begin{aligned}
L_1 &= -\{P_0, N_3\} + \{P_1, M_2\}\\
&\quad + \{P_2, M_1\} + P_3^2 - P_2^2 + (P_0 + P_1)^2,\\
L_2 &= -2P_1N_3 - 2P_0M_2 - 2\{P_1, M_2\}\\
&\quad + M_3^2 - N_1^2 - N_2^2 + P_2^2 - 2(P_0 + P_1)P_0,\\
L_3 &= \{P_0 + P_1, N_3 + M_2\} - N_2^2.
\end{aligned}
\tag{3.93}
$$

The coordinates x_i can vary in the ranges

$x_1 > 1 > x_2, x_3 > 0 > x_4$; $x_1 > 1 > 0 > x_2, x_3, x_4$;

$x_1, x_2, x_3 > 1 > x_4 > 0$.

(125) $(t, x, y, z) \to (ix, it, y, z)$

(126) This type corresponds to $f(x) = x^3$ and

$$
\begin{aligned}
&(x - t)^2 = -x_1x_2x_3x_4, \quad z = \tfrac{1}{2}(x_1 + x_2 + x_3 + x_4),\\
&2y(x - t) = x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4,\\
&t^2 - x^2 - y^2 = x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4.
\end{aligned}
\tag{3.94}
$$

The Klein–Gordon equation assumes the form (3.76) with $\partial/\partial\nu_j = 2x_j^{3/2}(\partial/\partial x_j)$.

The three operators are

$$L_1=\{P_1,M_2\}-\{P_0,N_3\}-\{P_2,M_1\},$$
$$L_2=\{P_2,N_3+M_2\}+\{M_2,P_0+P_1\}+M_3^2-N_1^2-N_2^2-2P_0P_1-P_0^2, \tag{3.95}$$
$$L_3=\{P_0+P_1,N_3+M_2\}-(N_2-M_3)^2.$$

The coordinates x_i vary in the ranges $x_1>0>x_2,x_3,x_4$ or $x_1,x_2,x_3>0>x_4$.

(127)–(128)

(127) This type corresponds to $f(x)=x(x-1)$ and

$$2(t-x)=\tfrac{1}{2}(x_1^2+x_2^2+x_3^2+x_4^2)-(x_1x_2+x_1x_3+x_1x_4+x_2x_3+x_2x_4+x_3x_4)+(x_1+x_2+x_3+x_4),$$
$$2(t+x)=x_1+x_2+x_3+x_4,$$
$$y^2=(x_1-1)(x_2-1)(x_3-1)(x_4-1),$$
$$z^2=-x_1x_2x_3x_4. \tag{3.96}$$

In terms of these coordinates the Klein–Gordon equation assumes the form (3.76) with $\partial/\partial\nu_j=2\sqrt{x_j(x_j-1)}(\partial/\partial x_j)$.

The three operators are

$$4L_1=\{N_3+M_2,P_3\}-\{N_2-M_3,P_2\}+\tfrac{1}{2}\{N_1,P_0+P_1\}-\tfrac{1}{4}(P_0-P_1)^2+P_2^2+\tfrac{1}{2}P_0(P_0+P_1),$$
$$4L_2=(N_3+M_2)^2+\{M_2,P_3\}-\tfrac{1}{2}\{N_1,P_0+P_1\}+(N_2-M_3)^2+\tfrac{1}{2}\{N_2+M_3,P_2\},$$
$$4L_3=-P_3^2-M_1^2-(N_3+M_2)^2+\{N_3,P_3\}. \tag{3.97}$$

The coordinates x_i vary in the ranges

$x_1,x_2>1>x_3>0>x_4$; $1>x_1>0>x_2,x_3,x_4$;

and $1>x_1,x_2,x_3>0>x_4$.

(128) $(t,x,y,z)\to(x,t,iy,iz)$

(129) [$\mathfrak{C}$]

This type corresponds to $f(x)=x^2$ and

$$2(iz_1-z_2)=x_1+x_2+x_3+x_4,$$
$$2(iz_1+z_2)=\tfrac{1}{2}(x_1^2+x_2^2+x_3^2+x_4^2)-(x_1x_2+x_1x_3+x_1x_4+x_2x_3+x_2x_4+x_3x_4),$$
$$(z_3-iz_4)^2=x_1x_2x_3x_4,$$
$$(z_3^2+z_4^2)=-(x_1x_2x_3+x_1x_2x_4+x_1x_3x_4+x_2x_3x_4). \tag{3.98}$$

The Klein–Gordon equation assumes the form (3.76) with $\partial/\partial\nu_j=2x_j(\partial/\partial x_j)$.

The three operators are

$$4L_1=\{I_{32}+iI_{13},P_3\},+\{I_{42}+iI_{14},P_4\}+\tfrac{1}{2}\{I_{12},P_1+iP_2\},+(P_3-iP_4)^2-\tfrac{1}{4}(P_2+iP_1)^2,$$
$$4L_2=\{I_{23}+iI_{13},P_3\},+\{I_{24}+iI_{14},P_4\}+(I_{24}+iI_{32})^2-(I_{13}+iI_{14})^2,$$
$$4L_3=I_{34}^2-[I_{31}+I_{42}+i(I_{32}+I_{14})]^2+\tfrac{1}{2}\{I_{13}+I_{42}+i(I_{32}+I_{41}),P_4+iP_3\}. \tag{3.99}$$

(130) This type corresponds to $f(x)=x$ and

$$2(x-t)=1-x_1-x_2-x_3-x_4,$$
$$2y+(x-t)^2=x_2x_3+x_2x_4+x_2x_1+x_3x_1+x_3x_4+x_4x_1-(x_1+x_2+x_3+x_4)+1,$$
$$2(x+t)+2y(x-t)=-(x_1x_2x_3+x_1x_2x_4+x_1x_3x_4+x_2x_3x_4)+(x_1x_2+x_1x_3+x_1x_4+x_2x_3+x_2x_4+x_3x_4)-(x_1+x_2+x_3+x_4)+1,$$
$$z^2=-x_1x_2x_3x_4. \tag{3.100}$$

The Klein–Gordon equation assumes the form (3.76) with $\partial/\partial\nu_j=2\sqrt{x_j}(\partial/\partial x_j)$.

We have not yet determined the three operators which describe this system. The coordinates vary in the ranges $x_1,x_2,x_3>0>x_4$ and $x_1>0>x_2,x_3,x_4$.

(131) [$\mathfrak{C}$]

This type corresponds to $f(x)=1$ and

$$2(z_1+iz_2)=-(x_1+x_2+x_3+x_4),$$
$$2(z_3+iz_4)+(z_1+iz_2)^2=x_1x_2+x_1x_3+x_1x_4+x_2x_3+x_2x_4+x_3x_4,$$
$$(z_3-iz_4)+2(z_1+iz_2)(z_3+iz_4)=-x_1x_2x_3-x_1x_2x_4-x_1x_3x_4-x_2x_3x_4,$$
$$(z_1-iz_2)+(z_1+iz_2)(z_3-iz_4)+(z_3+iz_4)^2=x_1x_2x_3x_4. \tag{3.101}$$

The Klein–Gordon equation assumes the form (3.76) with $\partial/\partial\nu_j=\partial/\partial x_j$.

We have not yet determined the operators which describe this system.

This comples the list of orthogonal coordinates for which the Klein–Gordon equation separates. As was mentioned earlier we have only given those systems which are genuinely new in that they have not been derived elsewhere before. For the wave equation ($\lambda=0$) we have found 125 such coordinate systems. In addition there are 34 radial coordinate systems corresponding to the group reduction $E(3,1)\supset SO(3,1)\supset\{L_1,L_2\}$ where $[L_1L_2]=0$ and L_1,L_2 are second order elements in the enveloping algebra of SO(3, 1). Similarly there are 55 coordinate systems belonging to reductions of the type $E(3,1)\supset E(2,1)\supset\{L_1,L_2\}$, 11 coordinate systems belonging to reductions of the type $E(3,1)\supset E(3)\supset\{L_1L_2\}$,[14] and 36 coordinate systems belonging to reductions of the type $E(3,1)\supset E(2)\otimes E(1,1)\supset L_1\otimes L_2$, where in this last case L_1 and L_2 are second order elements in the enveloping algebras of E(2) and E(1, 1), respectively. We have a total of 261 coordinate systems in which the Klein–Gordon equation admits separation of variables.

[1]E.G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space—time. 1. The Lorentz group," J. Math. Phys. **18**, 1 (1977).
[2]E.G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space—time. 2. The group SO(4, $\mathbb{C}$)," SIAM J. Math. Anal. (to appear).
[3]E.G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space—time. 3. Semisubgroup coordinates," J. Math. Phys. **18**, 271 (1977).
[4]L.P. Eisenhart, "Separable systems of Stackel," Ann. Math. **35**, 284—305 (1934).
[5]S. Schweber, *Relativistic Quantum Field Theory* (Harper, New York, 1961).
[6]P. Winternitz, "The Poincare group, its little groups and their applications in particle physics," Preprint RPP-T3, Rutherford Laboratory, England (1969).
[7]H. Joos, "Zur darstellungs theorie der inhomogenen Lorentz gruppe als Grundlage quantenmechanischer Kinematik," Fortschr. Phys. **10**, 65—146 (1962).
[8]E.P. Wigner, "On unitary representations of the inhomogeneous Lorentz group," Ann. Math. **40**, 149—204 (1939).
[9]E.G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space—time. 4," University of Waikato Research Report, No. 37 (1976).
[10]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977).
[11]J. Meixner and F.W. Schafke, *Mathieusche Funktionen und Spharoidfunktionen* (Springer-Verlag, Berlin, 1954).
[12]W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[13]E.G. Kalnins and W. Miller, Jr., "Lie theory and separation of variables. 9. Orthogonal R-separable coordinate systems for the wave equation $\Psi_{tt} - \Delta_2\Psi = 0$," J. Math. Phys. **17**, 331—355 (1976).
[14]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., "Symmetry and separation of variables for the Helmholtz and Laplace equations," Nagoya Math. J. **60**, 35—80 (1976).

Lie theory and the wave equation in space-time. 5. *R*-separable solutions of the wave equation $\psi_{tt}-\Delta_3\psi=0$

E. G. Kalnins

Department of Mathematics, University of Waikato, Hamilton, New Zealand

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 6 April 1977)

A detailed classification is made of all orthogonal coordinate systems for which the wave equation $\psi_{tt}-\Delta_3\psi=0$ admits an R-separable solution. Only those coordinate systems are given which are not conformally equivalent to coordinate systems that have already been found in previous articles. We find 106 coordinates to given a total of 368 conformally inequivalent orthogonal coordinates for which the wave equation admits an R separation of variables.

INTRODUCTION

In this article we continue our investigation of the orthogonal R-separable coordinate systems for which the wave equation in space–time,

$$\psi_{tt}-\Delta_3\psi=0, \qquad (*)$$

admits an R-separation of variables.[1–4] In a previous article[4] we have studied coordinate systems for which the Klein–Gordon equation

$$\psi_{tt}-\Delta_3\psi=\lambda\psi \qquad (**)$$

admits a separation of variables. Such coordinate systems also admit a separation of variables for the wave equation (*). In paper 4 of this series we found 262 conformally inequivalent coordinate systems of this type. It is the purpose of this article to give those coordinate systems for which (*) admits a strictly R-separable solution. By this we mean those coordinate systems for which (*) admits an R-separable solution and for which there is no conformally equivalent coordinate system such that (*) is simply separable. As with the treatment of the wave equation in two space dimensions[5] we classify all different types of orthogonal coordinate systems whose coordinate curves are cyclides or their degenerate forms.

The content of the paper is arranged as follows. In Sec. I we discuss the relevant details concerning coordinate systems whose coordinate curves are cyclides of most general type. This is a development of the methods in the fundamental book by Bôcher.[6] Also in this section we give the various differential forms corresponding to the coordinate systems of interest. In Sec. II we present the various coordinate systems together with the corresponding separation equations and triplet of mutually commuting operators $\{L_1, L_2, L_3\}$ which describe each such system.

I. *R*-SEPARABLE DIFFERENTIAL FORMS FOR THE WAVE EQUATION

In this section we classify all the orthogonal differential forms for which the wave equation (*) admits a strictly "R-separable" separation of variables. We recall that if ψ is a solution of (*) which is R-separable in terms of some new coordinates x_i, $(i=1, 2, 3, 4)$, then ψ can be written in the form

$$\psi=\exp[Q(x_1, x_2, x_3, x_4)]\phi, \qquad (1.1)$$

where the wave equation for the function ϕ is such that ϕ admits a separation of variables. The factor $\exp Q$ is called the modulation function and has a definite form for each R-separable coordinate system. In addition no part of the function Q should contain the sum of functions of only one of the variables x_i. For a strict R-separable system the modulation function Q should not be zero. In a previous paper[5] where we treated the wave equation in two space variables it was shown that only coordinate systems whose coordinate curves were degenerate forms of confocal cyclides of the most general type were strictly R-separable. All remaining R-separable coordinate systems could be transformed into coordinate systems for which the Klein–Gordon equation $(\partial_{tt}-\Delta_2)\psi=\lambda\psi$ also admits a separation of variables. This was done by a suitable transformation of the O(3,2) conformal symmetry group of $(\partial_{tt}-\Delta_2)\psi=0$. The same situation is true in the case of three spatial dimensions and it is accordingly the purpose of this section to discuss confocal families of cyclides of general type and their degenerate forms. We now briefly outline the properties of cyclides of this type and refer the reader for details to our previous paper[5] and the book by Bôcher.[6]

Families of confocal cyclides have their natural setting in pentaspherical space. This is a six-dimensional space of six homogeneous coordinates $y_1:y_2:y_3:y_4:y_5:y_6$ which are not all simultaneously zero and which are connected by the relation

$$y_1^2+y_2^2+y_3^2+y_4^2+y_5^2+y_6^2=0. \qquad (1.2)$$

The space–time coordinates are related to the homogenous coordinates via the relations

$$\begin{aligned} &y_1=i(p^2-q^2-r^2-s^2+w^2),\\ &y_2=p^2-q^2-r^2-s^2-w^2, \qquad (1.3)\\ &y_3=2pw,\quad y_4=2ipw,\quad y_5=2irw,\quad y_6=2isw, \end{aligned}$$

where $t=p/w$, $x=q/w$, $y=r/w$, $z=s/w$. A cyclide is then defined as the locus of points lying on the quadric surface

$$\Phi=\sum_{ij=1}^{6}a_{ij}y_iy_j=0,$$

with $a_{ij}=a_{ji}$ and $\det(a_{ij})\neq 0$. The classification of

 0022-2488/78/1906-1247$1.00

cyclides under the group of orthogonal transformations which preserves the form

$$\sum_{i=1}^{6} y_i^2$$

is then the problem of classifying the intersections of two quadratic forms in six-dimensional projective space. This is performed by the method of elementary divisors applied to the two quadratic forms.

(For the details of this classification see Ref. 5 and 6.) The equation describing the most general family of confocal cyclides in six-dimensional pentaspherical space is

$$\sum_{i=1}^{6} \frac{y_i^2}{\lambda - e_i} = 0, \quad \sum_{i=1}^{6} y_i^2 = 0. \tag{1.4}$$

Here λ is one of the new curvilinear coordinates and $e_i \neq e_j$, if $i \neq j$ $(i, j = 1, \ldots, 6)$. If we choose an orthogonal coordinate system in space–time whose coordinate curves have equations of the type (1.4), then the line element in terms of these new coordinates becomes

$$ds^2 = \frac{1}{4\sigma w^2}\left[\sum_{i=1}^{4} \frac{(x_i - x_j)(x_i - x_k)(x_i - x_l)}{f(x_i)}\, dx_i^2\right], \tag{1.5}$$

where

$$f(x_i) = \prod_{j=1}^{6} (x_i - e_j) \quad \text{and} \quad -1\sqrt{\sigma} = \sum_{i=1}^{6} e_i y_i^2.$$

The pentaspherical coordinates y_i are related to the curvilinear coordinates x_i via the equations

$$y_i = \frac{\phi(e_i)}{f'(e_i)}, \quad i = 1, \ldots, 6, \tag{1.6}$$

where $\phi(\lambda) = \Pi_{j=1}^{4} (\lambda - x_j)$. If we write the solution ψ of the wave equation as

$$\psi = (\sigma^{1/2} w^2)\Phi, \tag{1.7}$$

then Φ satisfies the differential equation

$$\sum_{j=1}^{4} \left[\left(\frac{1}{\phi'(x_j)}\right)\frac{\partial^2 \Phi}{\partial v_j^2} + 3x_j \Phi\right] - 2\left(\sum_{i=1}^{6} e_i\right)\Phi = 0, \tag{1.8}$$

where $2dv_j = dx_j/\sqrt{f(x_j)}$. This equation admits separable solutions for the function Φ, i.e.,

$$\Phi = \prod_{j=1}^{4} E_j(x_j).$$

Each of the functions E_j satisfies the differential equation

$$\frac{d^2 E_j}{dv_j^2} + \left[3x_j^4 - 2\left(\sum_{i=1}^{6} e_i\right)x_j^3 + Ax_j^2 + Bx_j + C\right]E_j = 0. \tag{1.9}$$

We now proceed to classify coordinate systems of this type by considering the expression inside the square brackets in (1.5) and finding out what ranges of the coordinates x_i permit this differential form to have overall negative signature. We must also consider degenerate forms of these general coordinate systems which result when some of the e_i become equal. In addition we should mention that two confocal families of cyclides of type (1.4) are equivalent under the action of real linear transformations of the pentaspherical coordinates y_i which preserve the quantity $\sum_{i=1}^{6} y_i^2$ if their parameters e_i, e_i' and coordinates x_i, x_i' are related by the equations

$$e_i = \frac{\alpha e_i' + \beta}{\gamma e_i' + \delta}, \quad x_i = \frac{\alpha x_i' + \beta}{\gamma x_i' + \delta}, \tag{1.10}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ and $\alpha\delta - \beta\gamma \neq 0$. We now give the classification of the strictly R-separable coordinate systems, in particular the differential forms.

[1] The first type of differential form corresponds to R-separable coordinate systems of the type (1.6) for which all the e_i are real. In addition the relations (1.10) can be used to standardize these quantities so that $e_1 = \infty$, $e_2 = a$, $e_3 = b$, $e_4 = c$, $e_5 = 1$, $e = 0$ with $a > b > c > 1$. The differential form then becomes

$$ds^2 = \left(\frac{-y_1^2}{4w^2}\right)\left[\sum_{i=1}^{6} \frac{(x_i - x_j)(x_i - x_k)(x_i - x_l)}{h(x_i)}\, dx_i^2\right], \tag{1.11}$$

where $h(x) = (x - a)(x - b)(x - c)(x - 1)x$. The ranges of variation of the variables x_i are

$$\begin{aligned}
&x_1, x_2, x_3 > a > x_4 > b,\\
&\quad x_1, x_2 > a > b > x_3 > c > x_4 > 1;\\
&x_1, x_2, x_3 > a > b > x_4 > c;\\
&x_1 > a > x_2, x_3 > b > c > x_4 > 1;\\
&x_1, x_2 > a > x_3 > b > x_4 > c,\\
&\quad x_1 > a > b > x_2 > c > x_3 > x_4 > 0.
\end{aligned} \tag{1.12}$$

[2] The differential forms of this type are as in (1.11) but with

$$b = a^* = \alpha - i\beta, \quad \alpha, \beta \in \mathbb{R}.$$

The ranges of variation of the variables x_i are

$$\begin{aligned}
&x_1, x_2, x_3 > c > x_4 > 1 > 0,\\
&x_1, x_2 > c > x_3 > 1 > x_4 > 0.
\end{aligned} \tag{1.13}$$

[3] In this case the quantities e_i can be taken to be

$$e_1 = \infty, \quad e_2^* = e_3 = \gamma + i\delta, \quad e_4^* = e_5 = \alpha + i\beta,$$
$$e_6 = 0, \quad \alpha, \beta, \gamma, \delta \in \mathbb{R}.$$

The differential form is given as in (1.11) with

$$h(x) = [(x - \gamma)^2 + \delta^2][(x - \alpha)^2 + \beta^2]x.$$

The range of variation of the variables x_i are then

$$x_1, x_2, x_3 > 0 > x_4; \quad x_1 > 0 > x_2, x_3, x_4. \tag{1.14}$$

The simplest type of degenerate differential forms corresponding to cyclides of general type (1.4) are obtained by allowing pairs of the quantities e_i to become equal. This is achieved by the prescription given by Bôcher,[6] e.g., if e_1 and e_2 become equal, then they do so according to the prescription

$$e_1 = e_2 + \epsilon, \quad x_1 = e_2 + \epsilon x_1', \tag{1.15}$$

where ϵ is a first order quantity. With this substitution and the subsequent use of the relations (1.10) to take $e_1 = \infty$ the differential form becomes

$$ds^2 = \left(-\frac{(y_1^2 + y_2^2)}{4w^2}\right)\left[\frac{dx_1'^2}{x_1'(x_1' - 1)} - \sum_{i=2}^{4} \frac{(x_i - x_j)(x_i - x_k)}{h(x_i)}\, dx_i^2\right], \tag{1.16}$$

where $h(x) = (x - a)(x - b)(x - c)(x - d)$. If we make the same substitution in (1.6) relating the pentaspherical space coordinates y_i^2, we obtain

$$y_1^2=1-x_1',\quad y_2^2=x_1',$$
$$y_3^2=\frac{(x_2-e_3)(x_3-e_3)(x_4-e_3)}{(e_3-e_4)(e_3-e_5)(e_3-e_6)},$$
$$y_4^2=\frac{(x_2-e_4)(x_3-e_4)(x_4-e_4)}{(e_4-e_3)(e_4-e_5)(e_4-e_6)},\tag{1.17}$$
$$y_5^2=\frac{(x_2-e_5)(x_3-e_5)(x_4-e_5)}{(e_5-e_3)(e_5-e_4)(e_5-e_6)},$$
$$y_6^2=\frac{(x_2-e_6)(x_3-e_6)(x_4-e_6)}{(e_6-e_3)(e_6-e_4)(e_6-e_5)}.$$

In addition we note that the coordinate curve for the coordinate x_1' has the equation

$$\frac{y_1^2}{x_1'-1}+\frac{y_2^2}{x_1'}=0.\tag{1.18}$$

From the form of the pentaspherical coordinates in (1.6) we see that the real linear transformations which preserve the quantity $\sum_{i=1}^6 y_i^2$ form a group isomorphic to O(4,2). In fact the representation of a point in space–time by the six pentaspherical coordinates is such that the generators $L_{ij}=y_i\partial_{y_j}-y_j\partial_{y_i}$ are directly related to the canonical generators of the conformal symmetry group of the wave equation.[3] More specifically we have the relations

$$L_{12}=\tfrac{1}{2}(K_0-P_0),\quad L_{13}=\frac{i}{2}(K_1-P_1),\quad L_{14}=\frac{i}{2}(K_2-P_2),$$
$$L_{15}=\frac{i}{2}(K_2-P_3),\quad L_{16}=iD,\quad L_{23}=iN_1,\quad L_{24}=iN_2,$$
$$L_{25}=iN_3,\quad L_{26}=\frac{i}{2}(P_0+K_0),\quad L_{34}=M_3,\quad L_{35}=M_2,\tag{1.19}$$
$$L_{36}=-\tfrac{1}{2}(P_1+K_1),\quad L_{45}=M_1,\quad L_{46}=-\tfrac{1}{2}(P_2+K_2),$$
$$L_{56}=-\tfrac{1}{2}(P_3+K_3).$$

Here we have used the notation of Ref. 3 for the generators of the conformal symmetry group.

Taking note of these relations we see that coordinate systems of the type given by (1.17) corresponds to the diagonalization of the generator $L_{12}=y_1\partial_{y_2}-y_2\partial_{y_1}$. This generator may correspond to a rotation or a hyperbolic rotation in pentaspherical space. If it corresponds to a hyperbolic rotation we may always use an O(4,2) group motion to ensure that $L_{12}=D$. The resulting coordinate system in space–time is then equivalent to one of the radial coordinate systems discussed in Ref. 5. Accordingly in classifying differential forms of type (1.16) we need only consider those for which $0<x_1'<1$.

[4] If we choose $a=b$; $c=1$, $d=0$, then we have the possibilities

$$a>x_2>b>x_3>1>x_4>0.$$
$$x_2>a>x_3,\ x_4>b;\ x_2>a,\ 1>x_3,\ x_4>0;$$
$$x_2>a>x_3>b>1>x_4>0;\ b>x_2>1>x_3,\ x_4>0.$$
$$x_2,x_3,x_4>a;\ b>x_2,x_3,x_4>1,\ 0>x_2,x_3,x_4,$$
$$x_2,x_3>a;\ b>x_4>1,\ 0>x_4$$
$$x_2>a;\ b>x_3,x_4>1,\ 0>x_3,x_4,\ b>x_3>1>0>x_4.$$
$$b>x_2x_3>1>0>x_4,\ b>x_2>1>0>x_3,x_4,$$
$$a>x_2,x_3>b;\ b>x_4>1,\ 0>x_4.$$
$$a>x_2>b>1>x_3>0>x_4.$$

[5] If $a=b^*=\alpha+i\beta$, $\alpha,\beta\in\mathbb{R}$ and $c=1$, $d=0$, then we have the possibilities

$$x_2,x_3,x_4>1;\ x_2,x_3>1>0>x_4,$$
$$x_2>1>0>x_3,x_4,\ 0>x_2,x_3,x_4,\tag{1.21}$$
$$x_2>1>x_3,x_4>0\text{ and }1>x_2,x_3>0>x_4.$$

[6] If we have $a=b^*$ as above and $c=d^*=\gamma+i\delta$, $\gamma,\delta\in\mathbb{R}$, then the variables x_2, x_3, x_4 can be any real numbers. If in addition we allow e_3 and e_4 to become equal according to the prescription of Bôcher,[4]

$$e_3=e_4+\epsilon,\quad x_2=e_4+\epsilon x_2'.\tag{1.22}$$

The differential form is then

$$ds^2=\left(\frac{-(y_1^2+y_2^2)}{4w^2}\right)\left[\frac{dx_1'^2}{x_1'(x_1'-1)}+\frac{(e_4-x_3)(e_4-x_4)}{(e_4-e_5)(e_4-e_6)}\right.$$
$$\left.\times\frac{dx_2'^2}{x_2'(1-x_2')}+(x_4-x_3)\left(\frac{dx_3^2}{P(x_3)}-\frac{dx_4^2}{P(x_4)}\right)\right],\tag{1.23}$$

where $P(x)=(x-e_4)(x-e_5)(x-e_6)$. For all such differential forms $0<x_2'<1$. Differential forms of this type fall into classes in which the quantities e_4, e_5, e_6 can be chosen to be 0, 1, or a.

[7] $e_4=0$, $e_5=1$, $e_6=a$; $a>1$. The variables x_3, x_4 vary in the ranges

$$0<x_3<1<x_4<a,\quad 1<x_3<a<x_4,\quad x_3<0<1<x_4<a.\tag{1.24}$$

[8] $e_4=1$, $e_5=0$, $e_6=a$; $a>1$;

$$1<x_3<a<x_4,\quad x_3<0<x_4<1,\tag{1.25}$$
$$x_3<0<1<x_4<a,\quad 0<x_3<1<a<x_4.$$

If we now allow e_5 and e_6 to become equal by the usual prescription, the differential form becomes, taking $e_5=1$ and $e_6=0$,

$$ds=\left(\frac{(y_1^2+y_2^2)}{4w^2}\right)\left[\frac{dx_1'^2}{x_1'(x_1'-1)}+(1-x_4)\frac{dx_2'^2}{x_2'(1-x_2')}\right.$$
$$\left.+x_4\frac{dx_3'^2}{x_3'(1-x_3')}+\frac{dx_4^2}{x_4(1-x_4)}\right].\tag{1.26}$$

There is only one differential form of this type.

[9] For this case all the variables x_i' $(i=1,2,3)$, x_4 lie in the interval $[0,1]$.

A further class of differential forms can be obtained by taking

$$e_4=e_6+a\epsilon,\quad e_5=e_6+\epsilon,\quad x_i=e_6+\epsilon x_i',\quad i=3,4.\tag{1.27}$$

If we also put $e_6=\infty$ in the resulting differential form we obtain

$$ds=\left(\frac{-(y_4^2+y_5^2+y_6^2)}{4w^2}\right)\left[(x_2-x_1)\left(\frac{dx_1^2}{P(x_1)}-\frac{dx_2^2}{P(x_2)}\right)\right.$$
$$\left.+(x_3'-x_4')\left(\frac{dx_3'^2}{Q(x_3')}-\frac{dx_4'^2}{Q(x_4')}\right)\right],\tag{1.28}$$

where $P(x)=(x-e_1)(x-e_2)(x-e_3)$ and $Q(x)=(x-a)\times(x-1)x$. This differential form corresponds to the reductions $O(4,2)\supset O(3)\otimes O(2,1)$ and $O(4,2)\supset O(2,1)\otimes O(2,1)$ when expressed in elliptic coordinates in the case of the two reductions

$$O(3)\supset L \quad \text{and} \quad O(2,1)\supset L'. \tag{1.29}$$

With the exception of the reduction $O(2,1)\supset O(1,1)$ which can be conformally transformed into a radial system we can in principle write down all the differential forms corresponding to the reductions of the type $O(4,2)\supset O(3)\otimes O(2,1)$ and $O(4,2)\supset O(2,1)\otimes O(2,1)$ by considering degenerate forms of the differential form (1.29), but we do not do this here.

The remaining distinct type of differential form of interest in this section is obtained by taking $x_2=e_6+\epsilon' x_2'$ and $e_3=e_6+\epsilon'$ subsequent to the substitutions (1.27) and then allowing $e_6\to\infty$. We then obtain the differential form

$$ds^2=\left(\frac{(y_3^2+y_4^2+y_5^2+y_6^2)}{4w^2}\right)\left[\frac{dx_1^2}{x_1(1-x_1)}+\frac{dx_2'^2}{x_2'(x_2'-1)}+x_2'(x_3'-x_4')\left(\frac{dx_3'^2}{Q(x_3')}-\frac{dx_4'^2}{Q(x_4')}\right)\right]. \tag{1.30}$$

[10] In each class we have that $0<x_1<1$, $0<x_2'<1$. The remaining variables vary in the ranges

$$0<x_3'<1<x_4'<a, \quad 1<x_3'<a<x_4',$$

$$x_3'<0<1<a<x_4', \quad x_3'<0<x_4'<1.$$

A further differential form can be obtained from (1.29) by taking $a=1+\epsilon''$, $x_3'=1+\epsilon'' x_3''$. This gives one new differential form,

$$ds^2=\left(\frac{-(y_1^2+y_2^2+y_3^2+y_4^2)}{4w^2}\right)\left[\frac{dx_1^2}{x_1(1-x_1)}+\frac{dx_2'^2}{x_2'(x_2'-1)}+x_2'\left((1-x_4')\frac{dx_3''^2}{x_3''(x_3''-1)}+\frac{dx_4'^2}{x_4'(x_4'-1)}\right)\right], \tag{1.31}$$

where all the variables lie between 0 and 1.

[11] This gives one additional different form.

We have thus shown in this section how to get all the orthogonal coordinate systems we expect by various limiting procedures applied to coordinate systems of most general cyclidic type. We have as yet not fully understood in what sense these procedures are complete.

II. *R*-SEPARABLE COORDINATES FOR THE WAVE EQUATION

In this section we give the coordinate systems corresponding to the differential forms in Sec. I together with the separation equations. We also present the triplet L_1, L_2, L_3 of mutually commuting second order symmetric operators in the enveloping algebra of $O(4,2)$ whose eigenvalues are the separation constants for each coordinate system presented. We now tabulate the coordinate systems of interest starting with the most general real cyclidic type of coordinates.

A. Coordinate systems of class I

(1)–(5)

(a) A suitable choice of coordinates is

$$t=\frac{1}{R}\left[-\frac{(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-1)a}\right]^{1/2},$$

$$x=\frac{1}{R}\left[\frac{(x_1-b)(x_2-b)(x_3-b)(x_4-b)}{(b-a)(b-c)(b-1)b}\right]^{1/2},$$

$$y=\frac{1}{R}\left[\frac{(x_1-c)(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-1)c}\right]^{1/2}, \tag{2.1}$$

$$z=\frac{1}{R}\left[\frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(1-a)(1-b)(1-c)}\right]^{1/2},$$

where

$$R=\left(1+\left[\frac{x_1x_2x_3x_4}{abc}\right]^{1/2}\right).$$

The solution of the wave equation then assumes the form $\psi=R\Phi$ where $\Phi=\Pi_{i=1}^4 E_i(x_i)$ typically. The separation equations for the functions E_i are

$$\frac{d^2E_j}{dx_j^2}+\frac{1}{2}\left(\frac{1}{x_j-a}+\frac{1}{x_j-b}+\frac{1}{x_j-c}+\frac{1}{x_j-1}+\frac{1}{x_j}\right)\frac{dE_j}{dx_j}+\frac{(-2x_j^3+l_1x_j^2+l_2x_j+l_3)}{4(x_j-a)(x_j-b)(x_j-c)(x_j-1)x_j}E_j=0. \tag{2.2}$$

The operators whose eigenvalues are the separation constants are

$$\begin{aligned}L_1&=\tfrac{1}{4}(a+b+c)(P_3+K_3)^2+\tfrac{1}{4}(a+b+1)(P_2+K_2)^2\\&\quad+\tfrac{1}{4}(a+c+1)(P_1+K_1)^2-\tfrac{1}{4}(b+c+1)(P_0+K_0)^2\\&\quad+(a+b)M_1^2+(a+c)M_2^2-(b+c)N_3^2\\&\quad-(c+1)N_1^2-(b+1)N_2^2+(a+1)M_3^2,\\L_2&=\tfrac{1}{4}(ac+bc+ab)(P_3+K_3)^2\\&\quad+\tfrac{1}{4}(ab+a+b)(P_2+K_2)^2+\tfrac{1}{4}(ac+a+c)(P_1+K_1)^2\\&\quad-\tfrac{1}{4}(bc+b+c)(P_0+K_0)^2+abM_1^2\\&\quad+acM_2^2-bcN_3^2-cN_1^2-bN_2^2\\&\quad+aM_3^2,\\L_3&=-\tfrac{1}{4}abc(P_3+K_3)^2-\tfrac{1}{4}ab(P_2+K_2)^2\\&\quad-\tfrac{1}{4}ac(P_1+K_1)^2+\tfrac{1}{4}bc(P_0+K_0)^2.\end{aligned} \tag{2.3}$$

The coordinates x_i vary in the ranges

$$x_1>a>b>x_2>c>x_3>1>x_4>0.$$

There are four more coordinate systems of this type. We list below the complex transformation of the space time coordinates which relates the coordinates of type (a) to the new system together with the new ranges of variation of the coordinates x_i. The separation equations for the $E_j(x_j)$ are the same in each case and the basis defining operators can be obtained by the substitution given. We now list the possibilities.

(b) $(t,x,y,z)\to(iz,x,y,it)$

$x_1, x_2>a>x_3>b>x_4>c,$

(c) $(t,x,y,z)\to(x,t,iy,iz)$

$x_1>a>x_2,\ x_3>b>c>x_4>1,$

$x_1, x_2, x_3>a>b>c>x_4>1.$

(d) $(t, x, z) \to (it, ix, iy, iz)$

$x_1, x_2 > a > b > x_3 > c > x_4 > 1.$

(e) $(t, x, y, z) \to (t, ix, y, iz)$

$x_1, x_2, x_3 > a > x_4 > b.$

(6)—(7)

(a) A suitable choice of coordinates is

$$t+ix=\frac{1}{R}\left[\frac{2(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-1)a}\right]^{1/2},$$
$$y=\frac{1}{R}\left[\frac{(x_1-c)(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-1)c}\right]^{1/2}, \qquad (2.4)$$
$$z=\frac{1}{R}\left[\frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(1-a)(1-b)(1-c)}\right]^{1/2}.$$

where

$$R=\left(1+\left[\frac{x_1x_2x_3x_4}{abc}\right]^{1/2}\right)$$

and $a=b^*=\alpha+i\beta$; $\alpha, \beta \in \mathbb{R}$.

The solution of the wave equation has the form $\psi=R\Phi$ where each of the E_j satisfy Eq. (2.2). The operators whose eigenvalues are the separation constants are

$$\begin{aligned}L=&\tfrac{1}{4}(2\alpha+c)(P_3+K_3)^2+\tfrac{1}{4}(2\alpha+1)(P_2+K_2)^2\\&+2\alpha M_1^2+\tfrac{1}{4}(\alpha+c+1)[(P_1+K_1)^2\\&-(P_0+K_0)^2]-\frac{\beta}{4}[(P_0+K_0)(P_1+K_1)\\&+(P_1+K_1)(P_0+K_0)]+(\alpha+c)(M_2^2-N_3^2)\\&+\beta(N_3M_2+M_2N_3)+(\alpha+1)(M_3^2-N_2^2)\\&+\beta(N_2M_3+M_3N_2)-(c+1)N_2^2\\L_2=&-\tfrac{1}{4}(2\alpha c+\alpha^2+\beta^2)(P_3+K_3)^2\\&-\tfrac{1}{4}(2\alpha+\alpha^2+\beta^2)(P_2+K_2)^2\\&-(\alpha^2+\beta^2)M_1^2+\tfrac{1}{4}(\alpha c+\alpha+c)[(P_0+K_0)^2\\&-(P_1+K_1)^2]+\tfrac{1}{4}\beta(c+1)[(P_1+K_1)(P_0+K_0)\\&+(P_0+K_0)(P_1+K_1)]+\alpha c(N_3^2-M_2^2)\\&-c\beta(M_2N_3+N_3M_2)+cN_1^2\\&+\alpha(N_2^2-M_3^2)-\beta(M_3N_2+N_2M_3), \qquad (2.5)\\L_3=&\tfrac{1}{4}(\alpha^2+\beta^2)[c(P_3+K_3)^2+(P_2+K_2)^2]\\&+\frac{\alpha c}{4}[(P_0+K_0)^2-(P_1+K_1)^2]\\&-\frac{c\beta}{4}[(P_1+K_1)(P_0+K_0)+(P_0+K_0)(P_1+K_1)].\end{aligned}$$

The coordinates x_i can vary in the ranges

$x_1, x_2 > c > x_3 > 1 > x_4 > 0.$

(b) $(t, x, y, z) \to (it, ix, iy, iz)$

where $x_1, x_2, x_3 > c > x_4 > 1 > 0.$

(8)

A suitable choice of coordinates is

$$t+iy=\left[\frac{2(x_1-c)(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-d)c}\right]^{1/2}R,$$
$$x=\mathrm{Im}\left[\frac{2(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-d)a}\right]^{1/2}R, \qquad (2.6)$$
$$z=[-x_1x_2x_3x_4/abcd]^{1/2}/R,$$

where

$$R=\left\{1+\mathrm{Re}\left[-\frac{2(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-d)a}\right]^{1/2}\right\}$$

and $a=b^*=\alpha+i\beta$, $c=d^*=\gamma+i\delta$, $\alpha, \beta, \gamma, \delta \in \mathbb{R}$.

The solution of the wave equation has the form $\psi=R\Phi$ where each of the E_j satisfies the equation

$$\frac{d^2E_j}{dx_j^2}+\frac{1}{2}\left(\frac{1}{x_j-a}+\frac{1}{x_j-b}+\frac{1}{x_j-c}+\frac{1}{x_j-d}+\frac{1}{x_j}\right)\frac{dE_j}{dx_j}$$
$$+\frac{(-2x_j^3+l_1x_j^2+l_2x_j+l_3)}{4(x_j-a)(x_j-b)(x_j-c)(x_j-d)x_j}E_j=0. \qquad (2.7)$$

The operators whose eigenvalues are the separation constants are

$$\begin{aligned}L_1=&(2\alpha+\gamma)(M_1^2-N_3^2)+\delta(M_1N_3+N_3M_1)\\&+(2\gamma+\alpha)[M_2^2-\tfrac{1}{4}(P_3-K_3)^2]\\&+\tfrac{1}{2}\beta[M_2(P_3-K_3)+(P_3-K_3)M_2]\\&+\tfrac{1}{2}\gamma(P_0-K_0)^2-2\alpha N_2^2\\&+(\alpha+\beta)[\tfrac{1}{4}(P_0-K_0)^2-\tfrac{1}{4}(P_2-K_2)^2+M_3^2-N_1^2]\\&+\tfrac{1}{2}\beta[N_1(P_0-K_0)+(P_0-K_0)N_1]\\&+\tfrac{1}{4}\delta[(P_0-K_0)(P_2-K_2)+(P_2-K_2)(P_0-K_0)]\\&-\delta(N_1M_3+M_3N_1)-\frac{\beta}{2}[M_3(P_2-K_2)\\&+(P_2-K_2)M_3],\\L_2=&(\alpha^2+\beta^2+2\alpha\gamma)(N_3^2-M_1^2)-2\alpha\delta(M_1N_3\\&+N_3M_1)+(\gamma^2+\delta^2+2\alpha\gamma)[\tfrac{1}{4}(P_3-K_3)^2\\&-M_2^2]-\gamma\beta[M_2(P_3-K_3)+(P_3-K_3)M_2]\\&+(\alpha^2+\beta^2)N_2^2+\tfrac{1}{4}(\gamma^2+\delta^2)(P_1-K_1)^2\\&+\alpha\gamma[\tfrac{1}{4}(P_2-K_2)^2-\tfrac{1}{4}(P_0-K_0)^2+N_1^2-M_3^2]\\&-\frac{\gamma\beta}{2}[(P_0-K_0)N_1+N_1(P_0-K_0)]\\&-\frac{\alpha\delta}{4}[(P_0-K_0)(P_2-K_2)+(P_2-K_2)(P_0-K_0)]\\&-\frac{\beta\delta}{2}[M_3(P_0-K_0)+(P_0-K_0)M_3]\\&-\frac{\beta\delta}{2}[N_1(P_2-K_2)+(P_2-K_2)N_1]\\&+\alpha\delta(N_1M_3+M_3N_1)\\&+\frac{\gamma\beta}{2}[(P_2-K_2)M_3+M_3(P_2-K_2)], \qquad (2.8)\\L_3=&(\alpha^2+\beta^2)[\gamma(N_3^2-M_1^2)-\delta(N_3M_1+M_1N_3)]\\&+(\gamma^2+\delta^2)\{\alpha(\tfrac{1}{4}(P_3-K_3)^2-M_2^2)\\&-\frac{\beta}{2}[(P_3-K_3)M_2+M_2(P_3-K_3)]\}.\end{aligned}$$

The variables x_i can vary in the ranges $x_1 > 0 > x_2$, x_3, x_4 and $x_1, x_2, x_3 > 0 > x_4$.

B. Coordinate systems of class II

Coordinate systems this type consist of all the coordinate systems in which the operator $\frac{1}{2}(P_0 - K_0)$ is diagonal.

As has been discussed in Ref. 7 the R-separable solutions of (*) then have the form $\psi = (Y_0 - \cos\psi) \times \exp(i(2F+1)\psi)\Phi(Y_0, Y_1, Y_2, Y_3)$ where $Y_0^2 + Y_1^2 + Y_2^2 + Y_3^2 = 1$ and the space–time coordinates are given by

$$t = \frac{\sin\psi}{Y_0 - \cos\psi}, \quad x = \frac{Y_1}{Y_0 - \cos\psi}, \quad y = \frac{Y_2}{Y_0 - \cos\psi}, \quad z = \frac{Y_3}{Y_0 - \cos\psi}, \tag{2.9}$$

$i(2F+1)$ is the eigenvalue of the operator $\frac{1}{2}(P_0 - K_0)$ and F is a positive integer or half integer. The function Φ satisfies the equation

$$(\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{14}^2 + \Gamma_{23}^2 + \Gamma_{24}^2 + \Gamma_{34}^2)\Phi = -4F(F+1)\Phi, \tag{2.10}$$

where $\Gamma_{12} = -\frac{1}{2}(P_1 + K_1)$, $\Gamma_{13} = -\frac{1}{2}(P_2 + K_2)$, $\Gamma_{14} = -\frac{1}{2}(P_3 + K_3)$, $\Gamma_{23} = M_3$, $M_{24} = -M_2$, and $\Gamma_{34} = M_1$. Here we are using the notation of Ref. 5. The problem of separation of variables for coordinate systems in which $\frac{1}{2}(P_0 - K_0)$ is diagonal reduces to the problem of separation of variables on the three-dimensional sphere S_3 in four space. Acting on the functions Φ the operators given above have the form

$$\begin{aligned} &\Gamma_{12} = Y_0\partial_1 - Y_1\partial_0, \quad \Gamma_{13} = Y_0\partial_2 - Y_2\partial_0, \\ &\Gamma_{14} = Y_0\partial_3 - Y_3\partial_0, \quad \Gamma_{23} = Y_1\partial_2 - Y_2\partial_1, \\ &\Gamma_{24} = Y_1\partial_3 - Y_3\partial_1, \quad \Gamma_{34} = Y_2\partial_3 - Y_3\partial_2. \end{aligned} \tag{2.11}$$

This problem has been solved by Olevski[7,8] and the six coordinate systems on S_3 for which (2.10) admits separation of variables have recently been investigated.[7] In the interests of a complete presentation we give here the six coordinate systems mentioned, the separation equations, the operators describing the separation, and some comment on the actual solutions is also made where possible.

(9) Ellipsoidal coordinates

A suitable choice of coordinates is

$$\begin{aligned} Y_0^2 &= -\frac{(x_1-a)(x_2-a)(x_3-a)}{(b-a)(1-a)a}, \\ Y_1^2 &= -\frac{(x_1-b)(x_2-b)(x_3-b)}{(a-b)(1-b)b}, \\ Y_2^2 &= -\frac{(x_1-1)(x_2-1)(x_3-1)}{(a-1)(b-1)}, \\ Y_3^2 &= \frac{x_1x_2x_3}{ab}, \end{aligned} \tag{2.12}$$

where $0 < x_3 < 1 < x_2 < b < x_1 < a$.

The separation equations for $\Phi = E_1(x_1)E_2(x_2)E_3(x_3)$ have the form

$$\frac{dE_i}{dx_i} + \frac{1}{2}\left[\frac{1}{x_i - a} + \frac{1}{x_i - b} + \frac{1}{x_i - 1} + \frac{1}{x_i}\right]\frac{dE_i}{dx_i} + \frac{[4F(F+1)x_i^2 + l_1x_i + l_2]}{4(x_i - a)(x_i - b)(x_i - 1)x_i}E_i = 0. \tag{2.13}$$

The operators whose eigenvalues are the separation constants l_1 and l_2 are

$$\begin{aligned} L_1 &= +\tfrac{1}{4}(P_1 + K_1)^2 + \frac{b}{4}(P_2 + K_2)^2 + \tfrac{1}{4}(b+1)(P_3 + K_3)^2 \\ &\quad + aM_3^2 + (a+1)M_2^2 - (a+b)M_1^2, \\ L_2 &= \tfrac{1}{4}b(P_3 + K_3)^2 - aM_2^2 - abM_1^2, \end{aligned} \tag{2.14}$$

(10) Elliptic cylindrical coordinates of type I

A suitable choice of coordinates is

$$\begin{aligned} Y_0 &= \left(\frac{x_1x_2}{a}\right)^{1/2}\cos\phi, \quad Y_1 = \left(\frac{x_1x_2}{a}\right)^{1/2}\sin\phi, \\ Y_2 &= \left(\frac{(x_1-a)(x_2-a)}{a(a-1)}\right), \quad Y_3 = \left(\frac{(x_1-1)(x_2-1)}{(1-a)}\right)^{1/2}, \end{aligned} \tag{2.15}$$

where $0 < x_1 < 1 < x_2 < a$.

The separation equations have the form for $\Phi = E_1(x_1)E_2(x_2)A(\phi)$,

$$\frac{d^2E_i}{dx_i^2} + \frac{1}{2}\left[\frac{1}{x_i - a} + \frac{1}{x_i - 1} + \frac{2}{x_i}\right]\frac{dE_i}{dx_i} + \frac{[4F(F+1)x_i^2 + l_1x_i + l_2]}{(x_i - a)(x_1 - 1)x_i^2}E_i = 0, \tag{2.16}$$

where $i = 1, 2$,

$$a\frac{d^2A}{d\phi^2} + l_2A = 0.$$

The operators whose eigenvalues are the separation constants l_1 and l_2 are

$$\begin{aligned} L_1 &= +M_3^2 + \tfrac{1}{4}(P_2 + K_2)^2 + a[M_2^2 + \tfrac{1}{4}(P_3 + K_3)^2] \\ &\quad + \tfrac{1}{4}(a+1)(P_1 + K_1)^2, \\ L_2 &= -\frac{a}{4}(P_1 + K_1)^2. \end{aligned}$$

An alternative choice of coordinates is obtained by taking $x_1 = \mathrm{sn}^2(\rho_1, k)$ and $x_2 = (1/k^2)\,\mathrm{dn}^2(\rho_2, k')$ where $a = 1/k^2$. We then have that

$$\begin{aligned} y_0 &= \mathrm{sn}\rho_1\,dn\rho_2\cos\phi, \quad y_1 = \mathrm{sn}\rho_1\,dn\rho_2\sin\phi, \\ y_2 &= dn\rho_1\,\mathrm{sn}\rho_2, \quad y_3 = \mathrm{cn}\rho_1\,\mathrm{cn}\rho_2, \end{aligned} \tag{2.18}$$

where $0 \le \rho_1 < 2K$ and $-K' < \rho_2 < K'$. [Note: $\mathrm{sn}(z,k)$ is a Jacobi elliptic function.] In terms of these coordinates the solution for Φ has the form

$$\Phi = (\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2)^m K_{Fn}^{Ps}(\mathrm{dn}\rho_2)K_{Fn}^{Ps}(k\,\mathrm{sn}\rho_1)\begin{cases}\cos m\phi, \\ \sin m\phi.\end{cases} \tag{2.19}$$

Here $K_{Fn}^{Ps}(z)$ is an associated Lamé polynomial as defined in Ref. 7.

(11) Elliptic cylindrical coordinates of type II

A suitable choice of coordinates is

$$Y_0=\left(\frac{(x_1-1)(x_2-1)}{(1-a)}\right)^{1/2}\cos\phi,$$
$$Y_1=\left(\frac{(x_1-1)(x_2-1)}{(1-a)}\right)^{1/2}\sin\phi, \tag{2.20}$$
$$Y_2=\left(\frac{x_1x_2}{a}\right)^{1/2},\quad Y_3=\left(\frac{(x_1-a)(x_2-a)}{a(a-1)}\right)^{1/2},$$

where $0<x_1<1<x_2<a$. The separation equations have the form for $\Phi=E_1(x_1)E_2(x_2)A(\phi)$,

$$\frac{d^2E_i}{dx_i^2}+\frac{1}{2}\left[\frac{1}{x_i-a}+\frac{2}{x_i-1}+\frac{1}{x_i}\right]\frac{dE_i}{dx_i}$$
$$+\frac{[4F(F+1)x_i^2+l_1x_i+l_2]}{4(x_i-a)(x_i-1)2x_i}E_i=0, \tag{2.21}$$

where $i=1, 2$,

$$(a-1)\frac{d^2A}{d\phi^2}+l_2A=0.$$

The operators whose eigenvalues are the separation constants l_1 and l_2 are

$$L_1=M_1^2+\frac{(a-1)}{4}(P_1+K_1)^2+a[M_3^2+\tfrac{1}{4}(P_2+K_2)^2],$$
$$L_2=\frac{(1-a)}{4}(P_1+K_1)^2. \tag{2.22}$$

These coordinates can also be written in terms of Jacobian elliptic functions by the same substitution as used for system (10). We then obtain

$$Y_0=\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2\cos\phi,\quad Y_1=\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2\sin\phi,$$
$$Y_3=\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2,\quad Y_4=\mathrm{dn}\rho_1\,\mathrm{sn}\rho_2. \tag{2.23}$$

In terms of these coordinates the solution for Φ has the form

$$\Phi=(\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2)^m K_{Fn}^{Ps}\left(-\frac{ik'}{k}\,\mathrm{cn}\rho_2\right)K_{Fn}^{Ps}(\mathrm{cn}\rho_1)\begin{bmatrix}\cos m\phi,\\ \sin m\phi.\end{bmatrix} \tag{2.24}$$

(12) Spheroelliptic coordinates

A suitable choice of coordinates is

$$Y_0=\sin\alpha\left(\frac{x_1x_2}{a}\right)^{1/2},\quad Y_1=\sin\alpha\left(\frac{(x_1-1)(x_2-1)}{(1-a)}\right)^{1/2},$$
$$Y_2=\sin\alpha\left(\frac{(x_1-a)(x_2-a)}{a(a-1)}\right)^{1/2},\quad Y_3=\cos\alpha, \tag{2.25}$$

where $0<x_1<1<x_2<a$, $0<\alpha<\pi$.

The coordinate system can also be written in terms of elliptic functions as with coordinate systems (10) and (11). This gives the parametrization,

$$Y_0=\sin\alpha\,\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2,\quad Y_1=\sin\alpha\,\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2,$$
$$Y_2=\sin\alpha\,\mathrm{dn}\rho_1\,\mathrm{sn}\rho_2,\quad Y_3=\cos\alpha. \tag{2.26}$$

A typical solution for Φ is of the form $A(\alpha)E_1(\rho_1)E_2(\rho_2)$ where

$$E_1(\rho_1)E_2(\rho_2)=F_{ln}^{pq}(-i\rho_1+iK+K',\rho_2) \tag{2.27}$$

a product of Lamé polynomials and

$$A(\alpha)=(\sin\alpha)^lC_{2F-l}^{l+1}(\cos\alpha).$$

[Here $C_\mu^\nu(z)$ is a Gegebauer polynomial.] The two operators characterizing this system are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2+\tfrac{1}{4}(P_2+K_2)^2+M_3^2,$$
$$L_2=\tfrac{1}{4}(P_1+K_1)^2+\frac{a}{4}(P_2+K_2)^2, \tag{2.28}$$

with eigenvalues $-l(l+1)$ and λ_n^{pq} respectively.

(13) Spherical coordinates

A suitable choice of coordinates is

$$Y_0=\sin\alpha\sin\beta\cos\phi,\quad Y_1=\sin\alpha\sin\beta\sin\phi,$$
$$Y_2=\sin\alpha\cos\beta,\quad Y_3=\cos\alpha, \tag{2.29}$$

where $0\leq\alpha,\ \beta\leq\pi,\ \ 0\leq\phi<2\pi$.

A typical solution for Φ of the form $A(\alpha)B(\beta)C(\phi)$ is

$$\Phi=(\sin\alpha)^lC_{2F-l}^{l+1}(\cos\alpha)P_l^m(\cos\beta)\exp(im\phi). \tag{2.30}$$

The two operators characterizing this system are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2+\tfrac{1}{4}(P_2+K_2)^2+M_3^2$$

and

$$L_2=\tfrac{1}{4}(P_2+K_1)^2, \tag{2.31}$$

with eigenvalues $-l(l+1)$ and $-m^2$ respectively.

(14) Cylindrical coordinates

A suitable choice of coordinates is

$$Y_0=\sin\alpha\cos\beta,\quad Y_2=\sin\alpha\sin\beta,$$
$$Y_3=\cos\alpha\cos\phi,\quad Y_3=\cos\alpha\sin\phi, \tag{2.32}$$

where $0<\alpha<\pi$ and $0<\beta,\ \phi<2\pi$.

A typical solution for $\Phi=A(\alpha)B(\beta)C(\phi)$ is

$$\Phi=\exp[im\phi+ip\beta](\sin\alpha)^{a+b}(\cos\alpha)^{2F-a-b}$$
$$\times{}_2F_1(b-F,\ a-F,\ a+b+1;\ -\tan^2\alpha), \tag{2.33}$$

where $m=a+b$, $p=a-b$. The two operators characterizing this system are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2\ \text{ and }\ L_2=M_1^2 \tag{2.34}$$

with eigenvalues $-p^2$ and $-m^2$ respectively.

C. Coordinate systems of class III

These are the analogs of the elliptical coordinates of type (9). The difference in this case is that coordinate systems of this type correspond to the diagonalization of M_3^2 rather than $\frac{1}{4}(P_0-K_0)^2$. We now list the possible types of coordinates.

(15a) A suitable choice of coordinates is

$$t=\frac{1}{R}\left(\frac{(x_2-a)(x_3-a)(x_4-a)}{(b-a)(a-1)a}\right)^{1/2},$$
$$x=\frac{1}{R}\cos\phi,\quad y=\frac{1}{R}\sin\phi, \tag{2.35}$$
$$z=\frac{1}{R}\left(\frac{(x_2-b)(x_3-b)(x_4-b)}{(b-a)(b-1)b}\right)^{1/2},$$

where

$$R=\left[\left(\frac{(x_2-1)(x_3-1)(x_4-1)}{(a-1)(b-1)}\right)^{1/2}+\left(\frac{x_2x_3x_4}{ab}\right)^{1/2}\right].$$

The typical solution of the wave equation is $\psi=R\Phi$ where $\theta=E_2(x_2)E_3(x_3)E(x_4)A(\phi)$. The separation equa-

tions are the same as for system (9) with $A(\phi)=\exp(i(2F+1)\phi)$. The variables x_2, x_3, x_4 vary in the ranges $x_2, x_3 > a > b > x_4 > 1$,

$$b > x_2 > 1 > x_3, x_4 > 0, \quad b > x_2, x_3, x_4 > 1,$$

$$b > x_2 > 1 > 0 > x_3, x_4, \quad a > x_2, x_3 > b > x_4 > 1.$$

The operators whose eigenvalues are the separation constants are

$$\begin{aligned} L_1 &= (a+b)D^2 - \tfrac{1}{4}(a+1)(P_3-K_3)^2 + \tfrac{1}{4}(b+1)(P_0-K_0)^2 \\ &\quad + \frac{a}{4}(P_3+K_3)^2 - \tfrac{1}{4}b(P_0+K_0)^2 - N_3^2, \\ L_2 &= abD^2 + \tfrac{1}{4}a(P_3-K_3)^2 + \tfrac{1}{4}b(P_0-K_0)^2, \end{aligned} \tag{2.36}$$

and of course $L_3 = M_3^2$.

There are five further coordinate systems of this type. In each case we choose the x and y coordinates to be of the form

$$x=\frac{1}{R}\cos\phi, \quad y=\frac{1}{R}\sin\phi, \quad \text{and the operator } L_3 = M_3^2.$$

The separation equations are the same as in system (9). For each of these five further coordinate systems we give the choice of R and the coordinates t and z together with the form of the operators L_1 and L_2.

(16b) The modulation function R is

$$R=\left[\left(\frac{(x_2-1)(x_3-1)(x_4-1)}{(a-1)(b-1)}\right)^{1/2}+\left(\frac{(x_2-b)(x_3-b)(x_4-b}{(a-b)(b-1)b}\right)\right]^{1/2} \tag{2.37}$$

and the coordinates t and z are given by

$$t=\frac{1}{R}\left(\frac{x_2x_3x_4}{ab}\right)^{1/2}, \quad z=\frac{1}{R}\left(\frac{(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-1)a}\right)^{1/2}. \tag{2.38}$$

The operators L_1 and L_2 are

$$\begin{aligned} L_1 &= \tfrac{1}{4}(a+b)(P_0+K_0)^2 - \tfrac{1}{4}(a+1)(P_0-K_0)^2 \\ &\quad + (b+1)N_3^2 + aD^2 - \tfrac{1}{4}b(P_3+K_3)^2 + \tfrac{1}{4}(P_3-K_3)^2, \\ L_2 &= -\tfrac{1}{4}ab(P_0+K_0)^2 + \tfrac{1}{4}a(P_0-K_0)^2 - bN_3^2. \end{aligned} \tag{2.39}$$

The ranges of variation of the coordinates x_2, x_3, and x_4 are

$$x_2 > a > x_3, x_4 > b, \quad x_2 > a > b > x_3, x_4 > 1;$$

$$x_2, x_3, x_4 > a, \quad b > x_2, x_3, x_4 > 1, \quad a > x_2, x_3 > b > x_4 > 1,$$

and

$$x_2 > a > b > 1 > 0 > x_3, x_4.$$

(17c) This coordinate system is related to (16b) via the transformation $(t, x, y, z) \to (it, ix, iy, iz)$ of the space–time coordinates. The variables x_2, x_3, x_4 vary in the ranges

$$x_2 > a > x_3 > b > 1 > x_4 > 0 \text{ and } x_2 > a > b > 1 > x_3x_4 > 0.$$

(18d) This coordinate system is related to (16b) via the transformation $(t, x, y, z) \to (z, it, iy, t)$ of the space–time coordinates. The variables x_2, x_3, and x_4 vary in the ranges $x_2, x_3 > a > b > 1 > 0 > x_4, b > x_2, x_3 > 1 > 0 > x_4$ and $a > x_2, x_3 > b > 1 > 0 > x_4$.

(19e) This coordinate system is related the (15a) via the transformation $(t, x, y, z) \to (z, ix, iy, t)$ of the space–time coordinates. The variables x_2, x_3, and x_4 vary in the ranges $x_2 > a > b > x_3$, $x_4 > 1$.

(20f) This coordinate system is related to (16b) via the transformation $(t, x, y, z) \to (iz, x, y, it)$ of the space–time coordinates. The variables x_2, x_3, and x_4 vary in the ranges $a > x_2 > b > 1 > x_3 > 0 > x_4$.

In addition to the six types of coordinate systems we have discussed in class III we will also include coordinate systems corresponding to the differential form of type (1.16).

(21) A suitable choice of coordinates is

$$\begin{aligned} (z+it) &= \frac{1}{R}\left[\frac{2(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-1)a}\right]^{1/2}, \\ x &= \frac{1}{R}\cos\phi, \quad y=\frac{1}{R}\sin\phi, \end{aligned} \tag{2.40}$$

where

$$R=\left[\left(\frac{(x_2-1)(x_3-1)(x_4-1)}{(a-1)(b-1)}\right)^{1/2}+\left(\frac{x_2x_3x_4}{ab}\right)^{1/2}\right]. \tag{2.41}$$

The separation equations are given by (2.13). The operators whose eigenvalues are l_1 and l_2 are

$$\begin{aligned} L_1 &= 2\alpha D^2 + \tfrac{1}{4}(\alpha+1)[(P_3-K_3)^2-(P_0-K_0)^2] \\ &\quad - \frac{\beta}{2}(P_0P_3+K_0K_3) + \tfrac{1}{4}\alpha[(P_3+K_3)^2-(P_0+K_0)^2] - N_3^2, \\ L_2 &= (\alpha^2+\beta^2)D^2 + \tfrac{1}{4}\alpha[(P_3-K_3)^2-(P_0-K_0)^2] \\ &\quad + \tfrac{1}{4}\beta[(P_3-K_3)(P_0-K_0)+(P_0-K_0)(P_3-K_3)]. \end{aligned} \tag{2.42}$$

The variables x_2, x_3, and x_4 vary in the ranges

$$x_2, x_3, x_4 > 1, \quad x_2 > 1 > x_3, x_4 > 0, \quad x_2 > 1 > 0 > x_3, x_4.$$

(22) Coordinate systems of this type can be obtained from those of type (21) via the transformation $(t, x, y, z) \to (it, ix, iy, iz)$. The variables x_2, x_3, and x_4 lie in the ranges $x_2, x_3 > 1 > 0 > x_4$, $0 > x_2, x_3, x_4$, and $1 < x_2, x_3 > 0 > x_4$.

(23) A suitable choice of coordinates is

$$\begin{aligned} (z+it) &= \frac{1}{R}\left[\frac{2(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-d)}\right]^{1/2}, \\ x &= \frac{1}{R}\cos\phi, \quad y=\frac{1}{R}\sin\phi, \end{aligned} \tag{2.43}$$

here $R = \mathrm{Re}\omega - \mathrm{Im}\omega$

and

$$\omega=\left[\frac{2(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-d)}\right]^{1/2}.$$

The separation equations in the variables x_2, x_3 and x_4 are

$$\begin{aligned} &\frac{d^2E_i}{dx_i^2} + \frac{1}{2}\left[\frac{1}{x_i-a}+\frac{1}{x_i-b}+\frac{1}{x_i-c}+\frac{1}{x_i-d}\right]\frac{dE_i}{dx_i} \\ &\quad + \frac{[4F(F+1)x_i^2+l_1x_i+l_2]}{4(x_i-a)(x_i-b)(x_i-c)(x_i-d)}E_i = 0. \end{aligned} \tag{2.44}$$

The operators whose eigenvalues are l_1 and l_2 are

$$L_1 = -2\alpha D^2 - 2\gamma N_3^2 + \tfrac{1}{2}(\alpha+\gamma)[P_3K_3 + K_3P_3 - P_0K_0 - K_0P_0] + \tfrac{1}{2}\delta[P_0^2 - P_3^2 + K_3^2 - K_0^2] - \tfrac{1}{2}\beta[P_0K_3 + K_3P_0 + P_3K_0 + K_0P_3],$$

$$L_2 = (\alpha^2+\beta^2)D^2 + (\gamma^2+\delta^2)N_3^2 + \tfrac{1}{2}\alpha\gamma[P_3K_3 + K_3P_3 - P_0K_0 - K_0P_0] + \tfrac{1}{2}\alpha\delta[P_0^2 - P_3^2 + K_3^2 - K_0^2] + \beta\delta(P_0P_3 - K_0K_3) - \tfrac{1}{2}\beta\gamma[P_3K_0 + K_0P_3 + P_0K_3 + K_3P_0];$$

the variables x_2, x_3, and x_4 can assume any real values.

(24) A suitable choice of coordinates is

$$t+z = \frac{2}{R}\,\mathrm{Im}\left[\frac{(x_1-a)(x_2-a)(x_3-a)}{(a-b)^2}\right]^{1/2},$$

$$t-z = \frac{1}{R}\,\mathrm{Im}\left[\frac{1}{(a-b)} - \frac{1}{2}\left\{\frac{1}{x_1-a} + \frac{1}{x_2-a} + \frac{1}{x_3-a}\right\}\right],$$

$$x = \frac{1}{R}\cos\phi, \quad y = \frac{1}{R}\sin\phi,$$

where

$$R = 2\,\mathrm{Re}\left[\frac{(x_1-a)(x_2-a)(x_3-a)}{(a-b)^2}\right]^{1/2}$$

The separation equations in the variables x_2, x_3 and x_4 are

$$\frac{d^2E_i}{dx_i} + \left[\frac{1}{x_i-a} + \frac{1}{x_i-b}\right]\frac{dE_i}{dx_i} + \frac{[4F(F+1)x_i^2 + l_1x_i + l_2]}{4(x_i-a)^2(x_i-b)^2}E_i = 0. \tag{2.47}$$

The operators whose eigenvalues are l_1 and l_2 are

$$L_1 = \alpha[\tfrac{1}{4}(P_3 - P_0 - K_3 - K_0)^2 - (D+N_1)^2] + \tfrac{1}{2}\beta[(P_3 - P_0 - K_3 - K_0)(D+N_1) + (D+N_1)(P_3 - P_0 - K_3 - K_0)] + \alpha[\tfrac{1}{4}(P_0 + P_3 + K_3 - K_0)^2 - \tfrac{1}{4}(P_0 - P_3 + K_0 + K_3)^2 - (N_1 - D)^2 + \tfrac{1}{4}(P_0 + P_3 + K_0 - K_3)^2] - \tfrac{1}{4}[(P_0+K_3)(P_3+P_0+K_0-K_3) + (P_3+P_0+K_0-K_3)(P_0+K_3)],$$

$$L_2 = -\tfrac{1}{4}(P_0+K_3)^2 + \tfrac{1}{2}(\alpha^2+\beta^2)[\tfrac{1}{4}(P_0+P_3+K_3-K_0)^2 - \tfrac{1}{4}(P_0-P_3+K_0+K_3)^2 - (N_1-D)^2 - \tfrac{1}{4}(P_0+P_3+K_0-K_3)^2] + \tfrac{1}{2}(\alpha^2-\beta^2)[\tfrac{1}{4}(P_3-P_0-K_0-K_3)^2 - (D+N_1)^2] - \frac{\alpha\beta}{4}[(P_3-P_0-K_3-K_0)(D+N_1) + (D+N_1)(P_3-P_0-K_3-K_0)] + \tfrac{1}{4}(P_0+K_3)\left[\beta(D-N_1) - \frac{\alpha}{2}(P_0+P_3+K_0-K_3)\right] + \tfrac{1}{4}\left[\beta(D-N_1) - \frac{\alpha}{2}(P_0+P_3+K_0-K_3)\right](P_0+K_3). \tag{2.48}$$

(25) This coordinate system is of similar type to coordinate systems (10) and (11) appearing in Class II. A suitable choice of coordinates is

$$t = \frac{1}{R}\left(\frac{(x_1-a)(x_2-a)}{a(a-1)}\right)^{1/2}, \quad x = \frac{1}{R}\cos\psi\left(\frac{(x_1-1)(x_2-1)}{(a-1)}\right)^{1/2},$$

$$y = \frac{1}{R}\cos\phi, \quad z = \frac{1}{R}\sin\phi,$$

where

$$R = \left(\frac{(x_1-1)(x_2-1)}{(a-1)}\right)^{1/2}\sin\psi + \left(\frac{x_1x_2}{a}\right)^{1/2},$$

and $x_1, x_2 < 0$, $0 < x_1, x_2 < 1$.

The solution ψ of the wave equation has the form $\psi = R\Phi$. The separation equations for $\Phi = E_1(x_1)E_2(x_2) \times A(\phi)B(\psi)$ are

$$\frac{d^2E_i}{dx_i^2} + \frac{1}{2}\left[\frac{1}{x_i-a} + \frac{2}{x_i-1} + \frac{1}{x_i}\right]\frac{dE_i}{dx_i} \times \frac{[4F(F+1)(x_i-1)^2 + l_1(x_i-1) + l_2]}{4(x_i-a)(x_i-1)^2x_i}E_i = 0, \tag{2.50}$$

where $i = 1, 2$,

$$\frac{d^2A}{d\phi^2} = -(2F+1)^2A, \quad (a-1)\frac{d^2B}{d\psi^2} = l_2B.$$

The operators whose eigenvalues are the separation constants are

$$L_1 = (a-1)[D^2 + \tfrac{1}{4}(P_1-K_1)^2] - [N_1^2 + \tfrac{1}{4}(P_0+K_0)^2] + \frac{(a-2)}{4}(P_1+K_1)^2, \tag{2.51a}$$

$$L_2 = \frac{(a-1)}{4}(P_1+K_1)^2, \quad L_3 = M_1^2.$$

(26) A suitable choice of coordinates is

$$t = \frac{1}{R}\left(\frac{(x_1-a)(x_2-a)}{(a-1)}\right)^{1/2}, \quad x = \frac{1}{R}\cos\psi\left(\frac{-x_1x_2}{a}\right)^{1/2}, \quad y = \frac{1}{R}\cos\phi, \quad z = \frac{1}{R}\sin\phi, \tag{2.51b}$$

where

$$R = \left[\left(\frac{-x_1x_2}{a}\right)^{1/2}\sin\psi + \left(\frac{(x_1-1)(x_2-1)}{(1-a)}\right)\right]^{1/2}$$

and $x_1 < 0 < 1 < x_2 < a$.

The solution ψ of the wave equation has the form $\psi = R\Phi$. The separation equations for $\Phi = E_1(x_1)E_2(x_2) \times A(\phi)B(\psi)$ are

$$\frac{d^2E_i}{dx_i^2} + \frac{1}{2}\left[\frac{1}{x_i-a} + \frac{1}{x_i-1} + \frac{2}{x_i}\right]\frac{dE_i}{dx_i} + \frac{[4F(F+1)x_i^2 + l_1x_i + l_2]}{4(x_i-a)(x_i-1)x_i^2}E_i \equiv 0 \tag{2.52}$$

where $i = 1, 2$,

$$\frac{d^2A}{d\phi^2} = -(2F+1)^2A, \quad a\frac{d^2B}{d\psi^2} = l_2B.$$

The operators whose eigenvalues are the separation constants are

$$L_1 = -a[D^2 + \tfrac{1}{4}(P_1-K_1)^2] - M_{01}^2 + \tfrac{1}{4}(P_2+K_2)^2 + \frac{(a+1)}{4}(P_1+K_1)^2, \tag{2.53}$$

$$L_2 = -\frac{a}{4}(P_1+K_1)^2, \quad L_3 = M_1^2.$$

This completes the list of coordinate systems of Class III.

D. Coordinate systems of class IV

Coordinate systems of this type correspond to the two direct product reductions $SO(4,2)\supset SO(2,1)\otimes SO(2,1)$ and $SO(4,2)\supset SO(3)\otimes SO(1,2)$. In each of these cases coordinates can be chosen from the nine separable classes of orthogonal coordinates on the two sheeted and one sheeted two-dimensional hyperboloids and the two separable classes of orthogonal coordinate systems on the two-dimensional sphere. The coordinate systems on these manifolds are given in the Appendix. In classifying coordinates of this type we give the general form of the space–time coordinates in terms of the above mentioned two-dimensional manifolds.

(1) Coordinate systems corresponding to the reduction

$$SO(4,2)\supset SO(3)\otimes SO(1,2)$$

A suitable choice of space–time coordinates is

$$t=\frac{\xi_2}{\xi_1+\xi_3},\quad x=\frac{\zeta_1}{\xi_1+\xi_3},\qquad y=\frac{\zeta_2}{\xi_1+\xi_3},\quad z=\frac{\zeta_3}{\xi_1+\xi_3}, \tag{2.54}$$

where $\xi_1^2-\xi_2^2-\xi_3^2=-1$ and $\zeta_1^2+\zeta_2^2+\zeta_3^2=1$.

With the exception of coordinate systems of type (8) (which can always be chosen such that D is diagonal) there are 16 coordinate systems of this type on the single and double sheeted hyperboloids. In each case the solution of the wave equation has the form

$$\psi=(\xi_1+\xi_3)\phi(\zeta_1,\zeta_2,\zeta_3)\theta(\xi_1,\xi_2,\xi_3)$$

where the functions ϕ and θ satisfy the equations

$$(M_1^2+M_2^2+M_3^2)\phi=-l(l+1)\phi,\qquad [\{P_0,K_0\}+D^2]\theta=l(l+1)\theta, \tag{2.55}$$

where l is a positive integer. The operators corresponding to each of the 16 possible coordinate systems can then be read off from the Appendix, if we make the identifications $N_1=\frac{1}{2}(P_0+K_0)$, $N_2=D$, and $M_3=\frac{1}{2}(P_0-K_0)$ in the case of the SO(1, 2) coordinates.

(2) Coordinate systems corresponding to the reduction

$$SO(4,2)\supset SO(2,1)\otimes SO(2,1).$$

A suitable choice of space–time coordinates is

$$t=\frac{\xi_1}{\xi_1+\xi_3},\quad x=\frac{\xi_2}{\xi_1+\xi_3},\qquad y=\frac{\zeta_2}{\xi_1+\xi_3},\quad z=\frac{\zeta_3}{\xi_1+\xi_3}, \tag{2.56}$$

where

$$\zeta_1^2-\zeta_2^2-\zeta_3^2=\epsilon,\quad \xi_1^2-\xi_2^2-\xi_3^2=-\epsilon,\quad \epsilon=\pm 1.$$

Again with the exception of coordinate systems of type (8) there are 64 coordinate systems of this type. In each case the solution of the wave equation has the form $\psi=(\xi_1+\xi_3)\phi(\xi_1,\xi_2,\xi_3)\theta(\zeta_1,\zeta_2,\zeta_3)$, where the functions ϕ and θ satisfy the equations

$$(N_2^2+N_3^2-M_1^2)\theta=j(j+1)\theta,\qquad [-\{P_1,K_1\}+D^2]\phi=j(j+1)\phi, \tag{2.57}$$

where

$$j=-\tfrac{1}{2}+iq,\quad 0<q<\infty.$$

The operator corresponding to the SO(2,1) associated with the vector (ξ_1,ξ_2,ξ_3) can be read off from the Appendix with the identifications $N_2=\frac{1}{2}(P_1-K_1)$, $N_2=D$, and $M_3=\frac{1}{2}(P_1+K_1)$.

We have looked at four classes of coordinate systems for which the wave equation (*) is strictly R-separable and found 106 distinct such coordinate systems. This gives a total of 368 inequivalent R-separable coordinate systems for the wave equation (*).

APPENDIX

In this appendix we list the orthogonal separable coordinate systems for the two-dimensional sphere, single sheeted, and double sheeted hyperboloids. In each case we list the symmetric second order operator in the enveloping algebra of the symmetry groups of these manifolds which describes each coordinate system. The coordinates (with the exception of the single sheet hyperboloid) can be found in the article by Olevski[8] and the operator characterization is due to Winternitz *et al.*[9]

A. Coordinate systems separable on the two-dimensional sphere $\zeta_1^2+\zeta_2^2+\zeta_3^2=1$

If we write the generators $M_1=\zeta_2\partial_{\zeta_3}-\zeta_3\partial_{\zeta_2}$, $M_2=\zeta_1\partial_{\zeta_3}-\zeta_3\partial_{\zeta_1}$, and $M_3=\zeta_1\partial_{\zeta_2}-\zeta_2\partial_{\zeta_1}$, the coordinate systems and operators are:

(1) $(\zeta_1^{(1)})^2=\dfrac{x_1x_2}{a}$, $(\zeta_2^{(1)})^2=\dfrac{(x_1-1)(1-x_2)}{(a-1)}$

$$(\zeta_3^{(1)})^2=\frac{(x_1-a)(x_2-a)}{a(a-1)}\quad 0<x_1<1<x_2<a.$$

The operator is $L=aM_2^2+M_3^2$.

(2) $\zeta^{(2)}=(\cos x_1,\ \sin x_1\cos x_2,\ \sin x_1\sin x_2)$.

The operator is $L=M_1^2$.

B. Coordinate systems on the one and two sheeted two-dimensional hyperboloids $\xi_1^2-\xi_2^2-\xi_3^2=\pm 1$

We adopt the notation $N_1=\xi_1\partial_{\xi_2}+\xi_2\partial_{\xi_1}$, $N_2=\xi_1\partial_{\xi_3}+\xi_3\partial_{\xi_1}$, and $M_3=\xi_2\partial_{\xi_3}-\xi_3\partial_{\xi_2}$.

(1) $(\xi_1^{(1)})^2=\dfrac{x_1x_2}{a}$, $(\xi_2^{(1)})^2=\dfrac{(x_1-1)(x_2-1)}{(a-1)}$

$$(\xi_3^{(1)})^2=\frac{(x_1-a)(a-x_2)}{a(a-1)},\quad 1<x_1<a<x_2,$$

$$\xi^{(1)}\cdot\xi^{(1)}=(\xi_1^{(1)})^2-(\xi_2^{(1)})^2-(\xi_3^{(1)})^2=1.$$

The coordinates on $\xi\cdot\xi=-1$ are obtained by the substitution $\xi^{(1)}\to i\xi^{(1)}$ and $x_1<0<1<x_2<a$.

The operator is $L=N_1^2+aN_2^2$.

(2) $(\xi_1^{(2)})^2=\dfrac{(x_1-1)(1-x_2)}{(a-1)}$, $(\xi_2^{(2)})^2=-\dfrac{x_1x_2}{a}$,

$$(\xi_3^{(2)})^2=\frac{(x_1-a)(a-x_2)}{a(a-1)},$$

$$x_1<0<1<a<x_2,\quad \xi^{(2)}\cdot\xi^{(2)}=1.$$

The coordinates on the single sheeted hyperboloid $\xi = -1$ are obtained via the substitution $\xi \to i\xi$ and $1 < x_1, x_2 < a$; $x_1, x_2 > a$.

The operator is $L = N_1^2 - aM_3^2$.

(3) $(\xi_1^{(3)} + i\xi_2^{(3)})^2 = 2(x_1 - a)(x_2 - a)/a(a - b)$,

$a = b^* = \alpha + i\beta$, $(\xi_3^{(3)})^2 = -x_1x_2/ab$,

$x_1 < 0 < x_2$, $\xi^{(3)} \cdot \xi^{(3)} = 1$.

The transformation $\xi \to i\xi$ and $x_1, x_2 > 0$.

The operator is $L = \alpha(M_3^2 - N_2^2) + \beta\{M_3, N_2\}$

(4) $\xi_1^{(4)} + \xi_2^{(4)} = \sqrt{-x_1x_2}$,

$\xi_1^{(4)} - \xi_2^{(4)} = \sqrt{-x_1/x_2} + \sqrt{-x_2/x_1} + \sqrt{-x_1/x_2}$

$\xi_3^{(4)} = \sqrt{(1 - x_1)(x_2 - 1)}$, $x_1 < 0 < 1 < x_2$,

$\xi^{(4)} \cdot \xi^{(4)} = 1$.

The coordinates on the single sheeted hyperboloid are obtained via the substitution $\xi \to i\xi$ with $x_1, x_2 > 1$, $0 < x_1, x_2 < 1$, $x_1, x_2 < 0$.

The operator is $L = N_1^2 - (N_2 + M_3)^2$.

(5) $\xi_1^{(5)} + \xi_2^{(5)} = \sqrt{x_1x_2}$,

$\xi_1^{(5)} - \xi_2^{(5)} = -(\sqrt{x_1/x_2} + \sqrt{x_2/x_1} + \sqrt{x_1x_2})$,

$\xi_3^{(5)} = \sqrt{(x_1 - 1)(x_2 - 1)}$, $0 < x_1 < 1 < x_2$,

$\xi^{(5)} \cdot \xi^{(5)} = 1$.

The coordinates on the single sheet hyperboloid are obtained via the substitution $\xi \to i\xi$ with $x_1 < 0 < x_2 < 1$.

The operator is $L = N_1^2 + (N_2 + M_3)^2$.

(6) $\xi_1^{(6)} + \xi_2^{(6)} = \sqrt{-x_1x_2}$,

$\xi_1^{(6)} - \xi_2^{(6)} = (x_1 - x_2)/[4(-x_1x_2)^{3/2}]$,

$\xi_3^{(6)} = \frac{1}{2}\left[\left(-\frac{x_2}{x_1}\right)^{1/2} - \left(-\frac{x_1}{x_2}\right)^{1/2}\right]$, $x_1 < 0 < x_2$,

$\xi^{(6)} \cdot \xi^{(6)} = 1$.

The coordinates on the single sheet hyperboloid are obtained via the substitution $\xi \to i\xi$ with $x_1, x_2 > 0$.

The operator is $L = \{N_1, N_2 - M_3\}$.

(7) $\xi_1^{(7)} + \xi_2^{(7)} = \sqrt{x_1}$, $\xi_1^{(7)} - \xi_2^{(7)} = \frac{1}{\sqrt{x_1}} + \sqrt{x_1}x_2^2$,

$\xi_3^{(7)} = x_2\sqrt{x_1}$, $x_1, x_2 > 0$,

$\xi^{(7)} \cdot \xi^{(7)} = 1$.

The coordinates on the single sheet hyperboloid are obtained via the substitution $\xi \to i\xi$ with $x_1 < 0 < x_2$.

The operator is $L = (N_2 + M_3)^2$.

(8) $\xi^{(8)} = (\cosh x_1, \cosh x_2, \cosh x_1 \sinh x_2, \sinh x_1)$,

$\xi^{(8)} \cdot \xi^{(8)} = 1$,

$\xi^{(\hat{8})} = (\sinh x_1 \cosh x_2, \sinh x_1 \sinh x_2, \cosh x_1)$

$\xi^{(\hat{8})} = (\sin x_1 \sinh x_2, \sin x_1 \cosh x_2, \cos x_1)$,

$\xi^{(\hat{8})} \cdot \xi^{(\hat{8})} = -1$.

The operator is $L = N_1^2$.

(9) $\xi^{(9)} = (\cosh x_1, \sinh x_1 \cos x_2, \sinh x_1 \sin x_2)$,

$\xi^{(9)} \cdot \xi^{(9)} = 1$,

$\hat{\xi}^{(9)} = (\sinh x_1, \cosh x_1 \cos x_2, \cosh x_1 \sin x_2)$,

$\hat{\xi}^{(9)} \quad \hat{\xi}^{(9)} = -1$.

The operator is $L = M_3^2$.

[1] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 1. The Lorentz group," J. Math. Phys. 18, 1 (1977).

[2] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 2. The group SO(4, C)," SIAM J. Math. Anal. (to appear).

[3] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 3. Semisubgroup coordinates," J. Math. Phys. 18, 271 (1977).

[4] E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space–time. 4. The Klein–Gordon equation and the Poincare group," J. Math. Phys. 19, 1233 (1978).

[5] E. G. Kalnins and W. Miller, Jr., "Lie theory and separation of variables and orthogonal R-separable coordinate systems for the wave equation $\psi_{tt} - \Delta_2\psi = 0$," J. Math. Phys. 17, 331 (1976).

[6] M. Bocher, *Ueber die Reihenentwickelungen der Potentialtheorie* (Teubner, Leipzig, 1894).

[7] E. G. Kalnins, W. Miller, and P. Wintenitz, "The group O(4), separation of variables, and the Hydrogen atom," SIAM J. Appl. Math. (to appear).

[8] M. P. Olevski, Mat. Sbornik. 27, 379 (1950).

[9] P. Winternitz, I. Lukac, and Ya. A. Smorodinski, "Quantum numbers of the Little groups of the Poincare group," Sov. J. Nucl. Phys. 7, 139 (1968).

A variational principle for transport theory[a)]

Madhoo Kanal and Harry E. Moses

University of Lowell, College of Pure and Applied Science, Center for Atmospheric Research, Lowell, Massachusetts 01854
(Received 14 November 1977)

A maximal variational principle is used to construct an infinite medium Green's function for treating the boundary value problems of the linear transport theory (neutron and radiative). For the neutrons we consider the one-speed case and correspondingly for the radiative transfer the monochromatic case. The scattering properties of the medium are presumed to be dependent on the relaxation length. Thus, for the neutrons the secondary production function depends on the neutron's relaxation length and for the radiative transfer the albedo for single scattering is dependent on the optical depth. These two functions are kept arbitrary so that a large class of problems can be covered. The basic principle involves a functional which is an absolute maximum when the trial function is an exact solution of an integral equation of the Fredholm type. The kernel of the integral equation is required to satisfy certain symmetry and boundedness properties. We also exhibit an interesting relation between the absolute maximal and Schwinger's stationary variational principles, which in general is neither a maximum nor a minimum.

1. INTRODUCTION

A variational principle for the solution of integral equations with probability kernels was used by Demarcus[1] to solve the problem of Knudsen flow. It involves a functional which is an absolute maximum when the trial function is an exact solution of the transport equation. This variational principle is a special case of one discussed, for example in Ref. 2, for which a functional is an absolute maximum when the trial function is an exact solution of an integral equation of the Fredholm type, such that the kernel is required to satisfy certain symmetry and boundedness properties. The more general variational principle has already been used for the Gel'fand–Levitan equation.[3,4] In the present paper we use the variational principle to provide insights into solutions of unsolved problems of transport theory such as radiative or neutron transfer in inhomogeneous media.

The abstract principle is discussed for example in Mikhlin's book[2] which we follow. It has been utilized by Demarcus[1] and much later by Stokes and Demarcus.[5] Our work differs from that of the others in the fundamental sense that we use the principle to construct the approximate Green's function and to estimate the spectral function of the transport operator for the trial function chosen. Greater flexibility is thus achieved with respect to boundary conditions.

2. THE VARIATIONAL PRINCIPLE

The basic principle underlying the variational approach (see Ref. 2) is, that given the Fredholm integral equation

$$f(x)=\phi(x)+\int_a^b dy\,K(x,y)f(y), \tag{1}$$

the functional

$$F[n]=\int_a^b dx\,n(x)[2\phi(x)+\int_a^b dy\,K(x,y)n(y)-n(x)] \tag{2}$$

is an absolute maximum when $n(x)=f(x)$, for the kernel $K(x,y)$ which satisfies the following conditions of symmetry and boundedness:

(a) the kernel $K(x,y)$ is symmetric in x,y,

(b) $K(x,y)\geq 0$ for $a\leq x\leq b$ and $a\leq y\leq b$,

(c) $\int_a^b dx K(x,y)<1$ for $a\leq x\leq b$.

For most transport problems it turns out that the integral equation that one needs to consider (for example for an infinite medium Green's function) is not Eq. (1) but

$$f(x,x_0)=\phi(x,x_0)+\int_a^b dx'\,K(x,x')f(x',x_0), \tag{3}$$

where x_0 is an additional parameter. A further, rather pleasant, property that is encountered is when the inhomogeneous term $\phi(x,x_0)$ is the same as $K(x,x_0)$. Then Eq. (3) is in a symmetric form. That is, when

$$\phi(x,x_0)=K(x,x_0),$$

Eq. (3) becomes

$$f(x,x_0)=K(x,x_0)+\int_a^b dx'\,K(x',x)f(x',x_0), \tag{4}$$

and the general functional to be considered is

$$F[n]=\int_a^b dx\,n(x,x_0)[2K(x,x_0)+\int_a^b dx'K(x,x')n(x',x_0)-n(x,x_0)], \tag{5}$$

which is an absolute maximum when $n(x,x_0)=f(x,x_0)$ (c.f. Refs. 2 and 3). The resulting value of the functional is then

$$F[f]=\int_a^b dx\,f(x,x_0)K(x,x_0). \tag{6}$$

Using Eq. (4) in Eq. (6) to eliminate the integral we get

$$F[f]=\lim_{x\to x_0}[f(x,x_0)-K(x,x_0)]. \tag{7}$$

Thus, for an exact solution of the integral equation (4), the absolute maximum value of the functional is the difference of boundary values of $f(x,x_0)$ and the kernel $K(x,x_0)$ as x approaches x_0 from either side of x_0. Two further points are worth noting. First, in most tran-

[a)]Research supported by the Office of Naval Research.

 0022-2488/78/1906-1258$1.00

sport problems of interest the density function ($f(x, x_0)$) and the scattering kernel ($K(x, x_0)$) are singular at $x = x_0$. However, the difference of the two functions at $x = x_0$ is bounded from above. Furthermore, as we shall see later, the value of that difference is related to the spectral density of the transport operator corresponding to the choice of the trial function [for example see Eq. (57) in Sec. 3]. It is precisely for such a relation and the maximal property of the functional that it is possible to obtain the best estimates of the complete spectrum of the transport operator and hence the Green's function.

Now we consider an application of this formulation to the problems of radiative transfer in a vertically inhomogeneous atmosphere where the albedo for single scattering is an arbitrary function of the optical depth. The same application is valid, aside from changing of names, for the neutron transport in inhomogeneous media if the corresponding albedo for single scattering (or secondary production) of neutrons does not exceed unity anywhere in the given medium, i.e., for the subcritical problem.

3. APPLICATION TO PROBLEMS OF RADIATIVE TRANSFER AND ONE-SPEED NEUTRON TRANSPORT

We shall consider the problems of radiative transfer. Hence we use the corresponding appropriate terminology. The equation of radiative transfer for a plane parallel vertically inhomogeneous isotropically scattering atmosphere is

$$\mu \frac{\partial I}{\partial x}(x, \mu) + I(x, \mu) = \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' I(x, \mu') + Q(x, \mu), \tag{8}$$

where μ is the cosine of the angle between the vertical axis and the direction of propagation of radiation, x is the optical depth, $\omega(x)$ is the albedo for single scattering, and $I(x, \mu)$ is the specific intensity. We have added a source term $Q(x, \mu)$ in case one is dealing with neutron transport. For problems of radiative transfer Q will be zero in general. Let the atmosphere be finite and optically bounded between $x = x_a$ and $x = x_b$. In order to set up the boundary value problem, we follow the standard procedure discussed in Refs. 6 and 7. Thus, consider a time reversed adjoint equation

$$\begin{aligned} -\mu \frac{\partial}{\partial x} G(x, -\mu \to x_0, -\mu_0) + G(x, -\mu \to x_0, -\mu_0) \\ = \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' G(x, -\mu' \to x_0, -\mu_0) \\ + \frac{\sqrt{\omega(x)}}{2} \delta(x - x_0)\delta(\mu - \mu_0). \end{aligned} \tag{9}$$

Note that, here, the delta functions are weighted by $\sqrt{\omega(x)}/2$. Multiplying Eq. (8) by G and Eq. (9) by I, subtracting and integrating the difference with respect to μ from -1 to $+1$, and x from x_a to x_b, we get the result

$$\begin{aligned} I(x, \mu) = \frac{2}{\sqrt{\omega(x)}} \int_{-1}^{1} d\mu' \mu' [I(x_a, \mu') G(x_a, -\mu' \to x, -\mu) \\ - I(x_b, \mu') G(x_b, -\mu' \to x, -\mu)] \\ + \frac{2}{\sqrt{\omega(x)}} \int_{x_a}^{x_b} dx' \int_{-1}^{1} d\mu' Q(x', \mu') \\ \times G(x', -\mu' \to x, -\mu). \end{aligned} \tag{10}$$

In particular, if

$$Q(x, \mu) = \frac{\sqrt{\omega(x)}}{2} \delta(x - x_1)\delta(\mu - \mu_1) \tag{10'}$$

and the medium is infinite, then $I(x, \mu) = \tilde{G}(x, \mu \to x_1, \mu_1)$ is a Green's function for the plane source at $x = x_1$ and $\mu = \mu_1$ satisfying Eq. (8). With suitable boundary conditions that both G and $\tilde{G}$ vanish at $x = \pm\infty$, we get from Eq. (10) the following reciprocity relation,

$$\sqrt{\omega(x)}\, G(x, -\mu \to x_1, -\mu_1) = \sqrt{\omega(x_1)}\, \tilde{G}(x_1, \mu_1 \to x, \mu). \tag{10''}$$

One may use this reciprocity relation to replace G by $\tilde{G}$. However, we prefer to use G instead. For the finite medium and any given incident conditions for $I(x, \mu)$ at the boundaries (i.e., at $x = x_a$, $\mu > 0$, $x = x_b$, $\mu < 0$), the reflected and transmitted radiations are determined from the limits of the integral identity (10). In other words, the unknown quantities $I(x_a, \mu)$ for $\mu < 0$ (the reflected intensity) and $I(x_b, \mu)$ for $\mu > 0$ (the transmitted intensity) may be determined from the integral equations

$$\begin{aligned} I(x_a, \mu) = \frac{2}{\sqrt{\omega(x_a)}} \int_{-1}^{1} d\mu' \mu' [I(x_a, \mu') G^{+}(x_a, -\mu' \to x_a, -\mu) \\ - I(x_b, \mu') G(x_b, -\mu' \to x_a, -\mu)] \\ + \frac{2}{\sqrt{\omega(x_a)}} \int_{x_a}^{x_b} dx' \int_{-1}^{1} d\mu' Q(x', \mu') G(x', -\mu' \to x_a, -\mu), \\ \mu > 0 \end{aligned} \tag{11}$$

and

$$\begin{aligned} I(x_b, \mu) = \frac{2}{\sqrt{\omega(x_b)}} \int_{-1}^{1} d\mu' \mu' [I(x_a, \mu') G(x_a, -\mu' \to x_b, -\mu) \\ - I(x_b, \mu') G^{-}(x_b, -\mu' \to x_b, -\mu)] \\ + \frac{2}{\sqrt{\omega(x_b)}} \int_{x_a}^{x_b} dx' \int_{-1}^{1} d\mu' Q(x', \mu') \\ \times G(x', -\mu' \to x_b, -\mu), \quad \mu < 0, \end{aligned} \tag{12}$$

where $+(-)$ on G represents its boundary value as x approaches the boundary of the atmosphere from the right (left). Solution of the integral equations (11) and (12) is possible provided we can construct the Green's function from the adjoint Eq. (9). In fact the proof of the existence of solutions of integral equations of this type for the grey problem is guaranteed by construction of the unknown coefficients (see Ref. 6 for further discussion). We focus our attention on procedure for the construction of G for the simple (and obvious) reason that it is the central quantity required for any boundary value problems of this kind. We shall demonstrate here the use of the variational technique to obtain an approximate solution to that problem of de-

termining G. To do that, we first obtain an integral equations for G. A convenient way is to take the Fourier transform of Eq. (9) with respect to x. The result is

$$G(x,-\mu\to x_0,-\mu_0)$$
$$=\frac{1}{4\pi}\int_{-\infty}^{\infty}dk\,\frac{\exp(ikx)}{1-ik\mu}\int_{-\infty}^{\infty}dx'\,\exp(-ikx')\omega(x')\int_{-1}^{1}d\mu'$$
$$\times\,G(x',-\mu'\to x_0,-\mu_0)+\sqrt{\omega(x_0)}\,\delta(\mu-\mu_0)$$
$$\times\frac{1}{4\pi}\int_{-\infty}^{\infty}dk\;\times\frac{\exp[ik(x-x_0)]}{1-ik\mu}. \tag{13}$$

For convenience in writing, denote the angular integral of the Green's function on the right-hand side of Eq. (13) by

$$H(x;x_0,\mu_0)=\int_{-1}^{1}d\mu\,G(x,-\mu\to x_0,-\mu_0). \tag{14}$$

Then we have

$$G(x,-\mu\to x_0,-\mu_0)$$
$$=\frac{1}{4\pi}\int_{-\infty}^{\infty}dk\,\frac{\exp(ikx)}{1-ik\mu}\int_{-\infty}^{\infty}dx'\,\omega(x')$$
$$\times\exp(-ikx')H(x';x_0,\mu_0)$$
$$+\sqrt{\omega(x_0)}\,\delta(\mu-\mu_0)\,\frac{1}{4\pi}\int_{-\infty}^{\infty}dk\,\frac{\exp[ik(x-x_0)]}{1-ik\mu}. \tag{15}$$

Clearly, if we can determine $H(x;x_0,\mu_0)$, then for any given $\omega(x)$, G is completely determined.

A. Integral equations for $H(x;x_0,\mu_0)$ and $\rho(x,x_0)$

By integrating Eq. (15) with respect to μ from -1 to $+1$ we obtain

$$H(x;x_0,\mu_0)=\frac{1}{\sqrt{\omega(x)}}\,\Psi(x;x_0,\mu_0)$$
$$+\frac{1}{\sqrt{\omega(x)}}\int_{-\infty}^{\infty}dx'\,K(x',x)\sqrt{\omega(x')}\,H(x';x_0,\mu_0), \tag{16}$$

where

$$\Psi(x;x_0,\mu_0)=\frac{1}{4\pi}\,\sqrt{\omega(x)\omega(x_0)}\int_{-\infty}^{\infty}dk\,\frac{\exp[ik(x-x_0)}{1-ik\mu_0} \tag{17}$$

and

$$K(x',x)=\frac{1}{4\pi}\,\sqrt{\omega(x')\omega(x)}\int_{-\infty}^{\infty}dk\,\exp[ik(x-x')]\int_{-1}^{1}\frac{d\mu}{1-ik\mu}$$
$$=\tfrac{1}{2}\sqrt{\omega(x')\omega(x)}\int_{0}^{1}\frac{d\mu}{\mu}\,\exp(-|x-x'|/\mu)$$
$$=\tfrac{1}{2}\sqrt{\omega(x')\omega(x)}\;E_1(|x-x'|), \tag{18}$$

where $E_1(|x-x'|)$ is the exponential integral. Note that Ψ and K are related by

$$K(x,x_0)=\int_{-1}^{1}d\mu\,\Psi(x;x_0,\mu). \tag{19}$$

To complete the set of integral equations that we need, we define the density $\rho(x,x_0)$ by

$$\rho(x,x_0)=\int_{-1}^{1}d\mu\,H(x;x_0,\mu) \tag{20}$$

and from (16) we find that $\rho(x,x_0)$ satisfies the following integral equation,

$$\sqrt{\omega(x)}\,\rho(x,x_0)=K(x,x_0)+\int_{-\infty}^{\infty}dx'K(x',x)$$
$$\times\sqrt{\omega(x')}\rho(x',x_0), \tag{21}$$

where we have used the integral relation (19) between Ψ and K. If we let

$$H_S(x;x_0,\mu_0)=\sqrt{\omega(x)}\;H(x;x_0,\mu_0) \tag{22}$$

and

$$\rho_S(x,x_0)=\sqrt{\omega(x)}\,\rho(x,x_0), \tag{23}$$

then the integral equations (16) and (21) assume symmetric looking forms:

$$H_S(x;x_0,\mu_0)=\Psi(x;x_0,\mu_0)$$
$$+\int_{-\infty}^{\infty}dx'K(x',x)H_S(x';x_0,\mu_0) \tag{24}$$

and

$$\rho_S(x,x_0)=K(x,x_0)+\int_{-\infty}^{\infty}dx'\,K(x',x)\rho_S(x',x_0). \tag{25}$$

B. Functionals for $H_S(x;x_0,\mu_0)$ and $\rho_S(x,x_0)$

Consider the functionals

$$F_1[N]=\int_{-\infty}^{\infty}dx\,N(x;x_0,\mu_0)[2\Psi(x;x_0,\mu_0)-N(x;x_0,\mu_0)$$
$$+\int_{-\infty}^{\infty}dx'\,K(x',x)N(x';x_0,\mu_0)] \tag{26}$$

and

$$F[n]=\int_{-\infty}^{\infty}dx\,n(x,x_0)[2K(x,x_0)-n(x,x_0)$$
$$+\int_{-\infty}^{\infty}dx'\,K(x',x)n(x',x_0)]. \tag{27}$$

Since the kernel $K(x',x)$ in both integral equations (24) and (25) satisfies the three properties listed below Eq. (2), we conclude that functionals $F_1[N]$[9] and $F[n]$ are absolute maxima for exact solutions of those integral equations (see Refs. 3 and 4 for the proof). Thus, for

$$N(x;x_0,\mu_0)=H_S(x;x_0,\mu_0)$$

and

$$n(x,x_0)=\rho_S(x,x_0),$$

the values of the corresponding functionals are

$$F_1[H_S]=\int_{-\infty}^{\infty}dx\,\Psi(x;x_0,\mu_0)H_S(x;x_0,\mu_0) \tag{28}$$

and

$$F[\rho_S]=\int_{-\infty}^{\infty}dx\,K(x,x_0)\rho_S(x,x_0). \tag{29}$$

Further use of Eq. (25) in Eq. (29) yields

$$F[\rho_S] = \lim_{x \to x_0} [\rho_S(x, x_0) - K(x, x_0)]. \tag{30}$$

Equation (30) is a rather surprising result for the reason that the value of the functional for the density bears such a simple relation to the density $\rho_S(x, x_0)$ and the scattering kernel $K(x, x_0)$. This is, in the strict sense, a relation between the direct and the inverse problem. For the Gel'fand—Levitan equation (c.f. Ref. 3) we discovered a relation identical to Eq. (30) where it turned out that the result can be translated into a theorem about the area under a curve to a given point x when considered as a functional of $n(x, x_0)$. This curve was given by the scattering potential to the given point when the functional took on its maximum value. In that case the functional may thus be considered as a method of obtaining the scattering potential from the spectral data through the variational technique. However, for the transport problems we are considering such a simple interpretation, is not possible. For one reason, we note that both $\rho_S(x, x_0)$ and $K(x, x_0)$ are singular at $x = x_0$. However, in the limit when $x \to x_0$ the difference of the boundary values of $\rho_S(x, x_0)$ and $K(x, x_0)$ is bounded. It would be of value to determine that limit.

C. Value of the functional $F[\rho_S]$

Clearly, it is of crucial importance to know the value of the functional $F[\rho_S]$ for the exact solution of the integral equation (27). The reason is simply that for any sequence of trial functions, the values of the corresponding functionals must approach the absolute maximum value obtained for the exact solution. One can then estimate the absolute error introduced for any particular choice of a trial function. Of course, in order to find $F[\rho_S]$, as given by Eq. (30), we must know the exact form of the density $\rho_S(x, x_0)$ (which in fact is the unknown). However, since the functional is the boundary value of the difference between $\rho_S(x, x_0)$ and $K(x, x_0)$ when x approaches x_0, we may take advantage of that to construct a pseudo Green's function, such that if $\rho_g(x, x_0)$ is the corresponding pseudo density, the difference $\rho_S(x, x_0) - \rho_g(x, x_0)$ in the limit $x \to x_0$ is small. In other words we require that $\rho_g(x, x_0)$ be singular in the same manner as $\rho_S(x, x_0)$ at $x = x_0$ plus a (possible) small correction term which is nonsingular at $x = x_0$ for all values of x_0. We proceed as follows: Consider Eq. (9) for the Green's function

$$\mu \frac{\partial}{\partial x} G(x, \mu \to x_0, \mu_0) + G(x, \mu \to x_0, \mu_0) = \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' \, G(x', \mu' \to x_0, \mu_0) + \frac{\sqrt{\omega(x)}}{2} \delta(x - x_0)\delta(\mu - \mu_0), \tag{31}$$

where for convenience we have changed the signs of μ and μ_0. Now consider a pseudo Green's function which satisfies

$$\mu \frac{\partial}{\partial x} G_g(x, \mu \to x_0, \mu_0) + G_g(x, \mu \to x_0, \mu_0) = \frac{\omega(x_0)}{2} \int_{-1}^{1} d\mu' \, G_g(x, \mu' \to x_0, \mu_0) + \frac{\sqrt{\omega(x)}}{2} \delta(x - x_0)\delta(\mu - \mu_0). \tag{32}$$

Note that ω associated with the first term on the right-hand side in Eq. (32) is a function of x_0 (the source point). If we let

$$L(x, \mu \to x_0, \mu_0) = G(x, \mu \to x_0, \mu_0) - G_g(x, \mu \to x_0, \mu_0), \tag{33}$$

then subtracting Eq. (32) from Eq. (31) we find that the difference $L(x, \mu \to x_0, \mu_0)$ satisfies the following equation,

$$\mu \frac{\partial}{\partial x} L(x, \mu \to x_0, \mu_0) + L(x, \mu \to x_0, \mu_0) = \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' \, L(x, \mu' \to x_0, \mu_0) + \frac{\omega(x) - \omega(x_0)}{2} \int_{-1}^{1} d\mu' \, G_g(x, \mu' \to x_0, \mu_0). \tag{34}$$

In what follows, we assume that $\omega(x)$ is either a continuous function or can be approximated by a suitable continuous function. If ω were independent of x, then $L(x, \mu \to x_0, \mu_0)$ would be the infinite medium Green's function without sources. In consequence, with appropriate boundary conditions at $x = \pm\infty$, the only solution of Eq. (34) would be the trivial one, i.e., $L = 0$, as one would expect. In that case G would coincide with G_g and the value of the functional [as given by Eq. (30)] may be readily shown to be proportional to the spectral density. For an arbitrary $\omega(x)$, the last term in Eq. (34) provides the "source" for $L(x, \mu \to x_0, \mu_0)$ and, therefore, that equation may have a nontrivial solution. However, we are only interested in the behavior of L near and at $x = x_0$. The first simple conclusion is that L is continuous everywhere on the real axis. This follows from the jump condition. Thus, if we integrate Eq. (34) from left to right of $x = x_0$, we get

$$\mu[L^+(x_0, \mu \to x_0, \mu_0) - L^-(x_0, \mu \to x_0, \mu_0)] = 0, \tag{35}$$

where we have assumed that $\omega(x)$ is continuous and

$$L^{\pm} = \lim_{\substack{x \to x_0 \\ \text{(Right of } x_0\text{)} \\ \text{(Left of } x_0\text{)}}} L(x, \mu \to x_0, \mu_0).$$

Continuity of L also implies that its derivative is also continuous at $x = x_0$. This follows from simply taking the difference of boundary values of Eq. (34) as $x \to x_0$ from both sides of x_0. If we solve Eq. (35) for μ, then the most general solution is

$$L^+(x_0, \mu \to x_0, \mu_0) - L^-(x_0, \mu \to x_0, \mu_0) = \lambda_0 \delta(\mu). \tag{36}$$

But, from Eq. (34) for $\mu = 0$, we get

$$L(x, 0 \to x_0, \mu_0) = \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' \, L(x, \mu' \to x_0, \mu_0) + \frac{\omega(x) - \omega(x_0)}{2} \int_{-1}^{1} d\mu' \, G_g(x, \mu' \to x_0, \mu_0). \tag{37}$$

Thus,

$$L^+(x_0, 0 \to x_0, \mu_0) - L^-(x_0, 0 \to x_0, \mu_0) = 0,$$

so that $\lambda_0 \equiv 0$ in Eq. (36). We have, therefore, shown that $L(x, \mu \to x_0, \mu_0)$ is continuous everywhere and so is its derivative.

We now proceed as before as we did with Eq. (9) and obtain the integral equation for $\rho_L(x,x_0)$ defined by

$$\rho_L(x,x_0)=\int_{-1}^{1} d\mu \int_{-1}^{1} d\mu_0\, L(x,\mu\to x_0,\mu_0). \tag{38}$$

The integral equation satisfied by $\rho_L(x,x_0)$ is

$$\rho_L(x,x_0)=\tfrac{1}{2}\int_{-\infty}^{\infty} dx'\, E_1(|x'-x|)\omega(x')\rho_L(x',x_0) + \tfrac{1}{2}\int_{-1}^{\infty} dx'\, E_1(|x'-x|)[\omega(x')-\omega(x_0)] \times \rho_g(x',x_0), \tag{39}$$

where

$$\rho_g(x,x_0)=\int_{-1}^{1} d\mu \int_{-1}^{1} d\mu_0\, G_g(x,\mu\to x_0,\mu_0) \tag{40}$$

is the pseudo density and $E_1(|x'-x|)$ is the exponential integral,

$$E_1(|x'-x|)=\int_0^1 \frac{d\mu}{\mu}\exp(-|x'-x|/\mu. \tag{41}$$

Since $\rho_L(x,x_0)$ is continuous at $x=x_0$, we get

$$\rho_L(x_0,x_0)=\tfrac{1}{2}\int_{-\infty}^{\infty} dx'\, E_1(|x'-x_0|)\omega(x')\rho_L(x',x_0) = \tfrac{1}{2}\int_{-\infty}^{\infty} dx'\, E_1(|x'-x_0|)[\omega(x')-\omega(x_0)]\rho_g(x',x_0). \tag{42}$$

Now we make the approximation for the first integral on the right-hand side of Eq. (42). The largest contribution that the exponential integral $E_1(|x'-x_0|)$ makes is near $x'=x_0$. In fact E_1 is weakly singular (logarithmically) there and decays rather rapidly away from x_0. In analogy with the method of stationary phases, we assume that $\rho_L(x',x_0)$ varies slowly near $x'=x_0$ and, therefore, may be taken out of the integral. Equation (42) then gives us

$$\rho_L(x_0,x_0)=\tfrac{1}{2}\{\int_{-\infty}^{\infty} dx'\, E_1(|x'-x_0|)[\omega(x')-\omega(x_0)] \times \rho_g(x',x_0)\}/[1-\tfrac{1}{2}\int_{-\infty}^{\infty} dx'\, E_1(|x'-x_0|)\omega(x')]. \tag{43}$$

But, by definition [see Eqs. (33), (38), and (40)]

$$\sqrt{\omega(x_0)}\,\rho_L(x_0,x_0) = \lim_{x\to x_0}[\sqrt{\omega(x)}\rho(x,x_0)-\sqrt{\omega(x_0)}\,\rho_g(x,x_0)]$$

therefore,

$$\sqrt{\omega(x_0)}\,\rho_L(x_0,x_0) = \lim_{x\to x_0}[\sqrt{\omega(x)}\,\rho(x,x_0)-K(x,x_0) - \sqrt{\omega(x_0)}\rho_g(x,x_0)+K(x,x_0)], \tag{44}$$

where $K(x,x_0)$ is, by definition (18), given by

$$K(x,x_0)=\tfrac{1}{2}\sqrt{\omega(x)\omega(x_0)}\,E_1(|x-x_0|). \tag{45}$$

Since, previously we defined $\rho_S(x,x_0)=\sqrt{\omega(x)}\rho(x,x_0)$ [see definition (23)] and now we define

$$\rho_{gS}(x,x_0)=\sqrt{\omega(x_0)}\,\rho_g(x,x_0), \tag{46}$$

we have from Eqs. (30) and (44) the value of the functional

$$F[\rho_S]=\lim_{x\to x_0}[\rho_S(x,x_0)-K(x,x_0)]$$

approximately equal to

$$F[\rho_S]\cong\lim_{x\to x_0}[\rho_{gS}(x,x_0)-K(x,x_0)] + \sqrt{\omega(x_0)}\,\rho_L(x_0,x_0), \tag{47}$$

where $\rho_L(x_0,x_0)$ is given by Eq. (43).

Here we have assumed that one can construct the Green's function $G_g(x,\mu\to x_0,\mu_0)$ by solving Eq. (32) and that the limit in Eq. (47) exists and is bounded. We answer that question by merely noting that the method of obtaining solution of Eq. (32) is in exact parallel to the case where the medium is homogeneous. In fact, following Case and Zweifel [see Eq. (11), p. 97, in Ref. 6], we find that

$$G_g(x,\mu\to x_0,\mu_0) = \pm\frac{\sqrt{\omega(x_0)}}{2}\left(\frac{\phi_{0\pm}(\mu)\phi_{0\pm}(\mu_0)\exp(-|x-x_0|/\nu_0)}{N_{0\pm}} + \int_0^1 \frac{d\nu}{N(\pm\nu)}\,\phi_{\pm\nu}(\mu)\phi_{\pm\nu}(\mu_0)\exp(-|x-x_0|/\nu)\right), \tag{48}$$

where $+(-)$ is for $x>x_0$ $(x<x_0)$, and $\phi_{0\pm}(\mu)$, $\phi_\nu(\mu)$ are discrete and continuum Case eigenfunctions given by

$$\phi_{0\pm}(\mu)=\frac{\nu_0\omega(x_0)}{2(\pm\nu_0-\mu)}, \tag{49}$$

$$\phi_\nu(\mu)=\frac{\nu\omega(x_0)}{2}\,P\frac{1}{\nu-\mu}+\lambda(\nu)\delta(\nu-\mu), \tag{50}$$

$$\lambda(\nu)=1-\nu\omega(x_0)\tanh^{-1}\nu, \tag{51}$$

$$N_{0+}=\frac{\omega(x_0)}{2}\,\nu_0^3\left(\frac{\omega(x_0)}{\nu_0^2-1}-\frac{1}{\nu_0^2}\right), \tag{52}$$

$$N(\nu)=\nu[(1-\omega(x_0)\nu\tanh^{-1}\nu)^2+\omega^2(x_0)\pi^2\nu^2/4], \tag{53}$$

and ν_0 is the zero of the dispersion function, i.e.,

$$1-\frac{\omega(x_0)\nu_0}{2}\ln\left(\frac{\nu_0+1}{\nu_0-1}\right)=0. \tag{54}$$

We remark here that Cases's discrete spectrum obtained from the solution of Eq. (54) is in fact now a "Regge" type trajectory determined by the functional form of $\omega(x_0)$. Further, the functional $F[\rho_S]$ will also follow a trajectory, which we will call "maximal trajectory." It is now a simple matter to find that maximal trajectory. This we do by noting that $\phi_{0\pm}(\mu)$ and $\phi_\nu(\mu)$ are normalized to unity, i.e.,

$$\int_{-1}^{1} d\mu\,\phi_{0\pm}(\mu)=\int_{-1}^{1} d\mu\,\phi_\nu(\mu)=1 \tag{55}$$

so that

$$\rho_{gS}(x,x_0)=\frac{\omega(x_0)}{2}\left(\frac{\exp(-|x-x_0|/\nu_0)}{N_{0+}} + \int_0^1 \frac{d\nu}{N(\nu)}\exp(-|x-x_0|/\nu)\right). \tag{56}$$

Hence

$$F[\rho_S] \cong \frac{\omega(x_0)}{2}\left[\frac{1}{N_{0+}} + \lim_{x\to 0}\int_0^1 d\nu \exp(-|x|/\nu)\left(\frac{1}{N(\nu)} - \frac{1}{\nu}\right)\right] + \sqrt{\omega(x_0)}\,\rho_L(x_0, x_0), \tag{57}$$

where, as a reminder, $\rho_L(x_0, x_0)$ is given by Eq. (43). From Eq. (57) we see that the functional is bounded. In the lowest approximation the continuum part, involving the integral, may be neglected; because, for small x the largest contribution from the integral involving $N(\nu)$ comes from near $\nu = 0$, where $N(\nu) \sim \nu$. Hence

$$F_0[\rho_S] = \frac{\omega(x_0)}{2N_{0+}} + \sqrt{\omega(x_0)}\,\rho_L(x_0, x_0), \tag{58}$$

where we have subscripted the functional with zero to indicate the zeroth approximation. We close this section with one more remark. Here we have demonstrated the use of a maximal variational principle to construct an appropriate Green's function for inhomogeneous media in dealing with problems of radiative and neutron transport. We have made no attempt to solve any actual problem for a real medium. We hope to make that a topic for the next paper. However, we present a simple example in Sec. 14, where we consider a trial function involving G_g.

We also wish to point out that the same variational principle seems to be applicable to wider variety of problems in theoretical physics, viz., kinetic theory of gases, plasma physics (such as the Vlasov equation), and electron transport in the upper atmosphere.

4. EXAMPLE

For the trial Green's function we take the simplest form which corresponds to the *scaling* of G_g, i.e., we let

$$G_T(x, \mu \to x_0, \mu_0) = \alpha\,\frac{\omega(x_0)}{\omega(x)}\,G_g(x, \mu \to x_0, \mu_0), \tag{59}$$

where α is a parameter to be determined from maximization of the functional (27). Now it follows that the trial density function is

$$\rho_T(x, x_0) = \alpha\,\frac{\omega(x_0)}{\omega(x)}\,\rho_g(x, x_0). \tag{60}$$

It is seen, by replacing $\omega(x)$ by $\omega(x_0)$ everywhere in Eq. (25), that $\rho_{gS} = \sqrt{\omega(x_0)}\,\rho_g$ satisfies the following integral equation

$$\rho_{gS}(x, x_0) = K_g(x, x_0) + \int_{-\infty}^{\infty} dx'\,K_g(x', x)\rho_{gS}(x', x_0), \tag{61}$$

where the kernel K_g is

$$K_g(x, x_0) = \frac{\omega(x_0)}{2}\,E_1(|x - x_0|). \tag{62}$$

Now in the expression (27) for the density functional we set $n = \rho_{TS} = \sqrt{\omega(x)}\,\rho_T(x, x_0)$ and compute the value of the functional. One may readily show via use of integral equation (61) that the value of the functional is

$$F[\rho_{TS}] = \alpha(2 - \alpha)\tilde{F}[\rho_{gS}] - \alpha^2 W(x_0), \tag{63}$$

where by definition

$$\tilde{F}[\rho_{gS}] = \int_{-\infty}^{\infty} dx\,K_g(x, x_0)\rho_{gS}(x, x_0) \tag{64}$$

and

$$W(x_0) = \int_{-\infty}^{\infty} dx\left(\frac{\omega(x_0)}{\omega(x)} - 1\right)\rho^2_{gS}(x, x_0). \tag{65}$$

Setting

$$\frac{\partial F}{\partial \alpha}[\rho_{TS}] = 0$$

we get

$$\alpha(x_0) = \frac{1}{1 + W(x_0)/\tilde{F}[\rho_{gS}]}. \tag{66}$$

Note that $\tilde{F}[\rho_{gS}]$ is the exact maximum value of the functional for the pseudo grey medium. The value of the functional $F[\rho_{TS}]$ is given by

$$F[\rho_{TS}] = \frac{\tilde{F}^2[\rho_{gS}]}{\tilde{F}[\rho_{gs}] + W(x_0)}. \tag{67}$$

To test the accuracy of the trial function chosen, we compute

$$\epsilon(x_0) = \frac{F[\rho_{TS}] - \tilde{F}[\rho_{gS}]}{F[\rho_{TS}]} \tag{68}$$

and see how far $\epsilon(x_0)$ is away from zero. Thus, for the present example considered $\epsilon(x_0)$ represents the local deviation from the grey approximation.

5. A RELATION BETWEEN THE MAXIMAL AND SCHWINGER'S VARIATIONAL PRINCIPLES

In the previous sections we have illustrated the basic ideas underlying the maximal variational principle as applied to transport theory. Because the functional used is an absolute maximum for the exact solution of the appropriate transport equation, it has the advantage that, in principle at least, one can obtain the estimate of absolute errors for any set of trial functions used. We have already provided an approximate uniform bound to the value of the functional which for most situations should be useful. By contrast other variational principles, such as Schwinger's or Rayleigh Ritz principles, which require stationarity only, the estimate of errors for a set of trial functions for the entire medium (i.e., the uniform error) is usually not feasible. In this sense those variational principles are rather limited and often very cumbersome when applied to real problems. However, it would be of pedagogical value to show a relation between the maximal principle which we have used and Schwinger's stationary principle.

Schwinger's variational principle[8] states that, given the Fredholm integral equation

$$y(r, r_0) = K(r, r_0) + \int_0^{\infty} dr'\,K(r', r)y(r'), \tag{69}$$

where the kernel $K(r', r)$ is symmetric, then the functional

$$S[n] = \{\int_0^{\infty} dr\,n(r, r_0)[n(r, r_0) - \int_0^{\infty} dr'\,K(r', r)n(r', r_0)]\}/ [\int_0^{\infty} dr\,n(r, r_0)K(r, r_0)]^2 \tag{70}$$

is *stationary* when $n(r,r_0)=y(r,r_0)$. The value of the functional is then

$$S[y]=\frac{1}{\int_0^\infty dr\,K(r,r_0)y(r,r_0)}. \tag{71}$$

Let

$$N[r_0]=\int_0^\infty dr\,K(r,r_0)n(r,r_0). \tag{72}$$

By comparing Schwinger's functional (70) and the maximal functional $F[n]$ (5), we conclude that the two functionals bear the following relationship with each other,

$$F[n]=2N(r_0)-S[n]N^2(r_0). \tag{73}$$

For the exact solution of Eq. (69), we have already from Eq. (70) that $S[y]=1/N(r_0)$. Therefore,

$$F[y]S[y]=1. \tag{74}$$

In other words, for the exact solution, the product of maximal functional and Schwinger's stationary functional is unity.

ACKNOWLEDGMENT

We are grateful to Professor K. M. Case of Rockefeller University for reading the manuscript and for making helpful suggestions.

[1]Wendell C. Demarcus, "The Problem of Knudsen Flow—Part II. Solution of Integral Equations with Probability Kernels," U.S.A.E.C. Report K-1302, Part II, Technical Information Office, Oak Ridge, Tennessee (1956).
[2]S.G. Mikhlin, *Variational Methods in Mathematical Physics*, Transl. by T. Boddington, edited by L.I.G. Chambers (MacMillan, New York, 1964).
[3]M. Kanal and H.E. Moses, "A Variational Principle for the Gelfand—Levitan Equation and the Korteweg—deVries Equation," J. Math. Phys. 12, 2445 (1977).
[4]I.M. Gelfand and B.M. Levitan, Am. Math. Soc. Trans. 1, 253 (1951).
[5]R.A. Stokes and W.C. Demarcus, Icarus, 14, 307—11 (1971).
[6]K.M. Case and P.F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[7]M. Kanal, J. Math. Phys. 11, 3042—52 (1970).
[8]J. Schwinger, Phys. Rev. 72, 742A (1947).
[9]Here $F_1[N]$ will in general be a distribution.

A new formula for the spinor norm

M. Perroud

Department of Mathematics, Ecole Polytechnique de Montréal, Montréal, Québec H3C 3A7, Canada
(Received 22 November 1977)

A formula is established, valid for all real pseudo-orthogonal groups, permitting an easy computation of the spinor norm of any pseudo-orthogonal matrix. It is a generalization of the well-known criterion used for O(3,1) according to which the spinor norm is determined by the sign of the last element of a Lorentz matrix.

The following notations are used: $O(p, q)$ denotes the pseudo-orthogonal group

$$O(p, q) = \{M \mid M \in \mathbb{R}^{n\times n};\ M^T G M = G\},$$

where

$$G = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix};$$

I_p and I_q denote the identity matrices of degree p and q, $n = p + q$, $pq \neq 0$.

When considering $\mathbb{R}^n$ as the direct sum $\mathbb{R}^p \oplus \mathbb{R}^q$, each vector X can be written uniquely as $X = x_1 + x_2$ with $x_1 \in \mathbb{R}^p$ and $x_2 \in \mathbb{R}^q$. The group $O(p, q)$ acts linearly on $\mathbb{R}^n$ and leaves invariant the bilinear form

$$(X, Y) = (x_1, x_1)_1 - (x_2, x_2)_2, \quad \text{where } (x_i, x_i)_i = x_i^T x_i.$$

With respect to this direct sum decomposition, each matrix $M \in O(p, q)$ can be written in the block form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{1}$$

The conditions $M^T G M = G$ and $M G M^T = G$ (the latter following from $M^{-1} = G M^T G$) impose for the blocks the conditions

$$\begin{aligned} A^T A &= I_p + C^T C, \\ A^T B &= C^T D, \qquad (2) \\ D^T D &= I_q + B^T B, \end{aligned}$$

$$\begin{aligned} A A^T &= I_q + B B^T, \\ A C^T &= B D^T, \qquad (3) \\ D D^T &= I_q + C C^T. \end{aligned}$$

All the above is equally valid if $\mathbb{R}$ is replaced by another field $\mathbb{F}$.

Lipschitz discovered in the nineteenth century[1] that the real pseudo-orthogonal groups contain normal subgroups of index 2 other than the special orthogonal group. Just as the existence of the latter may be established by means of a homomorphism of $O(p, q)$ onto a group of two elements (the determinant homomorphism), so the existence of these other subgroups may be established by means of a similar homomorphism, the so-called spinor norm. It is nowadays well known that this property also holds for pseudo-orthogonal groups defined over a broad class of fields $\mathbb{F}$. The spinor norm homomorphism is usually introduced through the Clifford groups,[2] but a direct definition exists which rests on the property, already known by Lipschitz for the real cases, that each pseudo-orthogonal transformation on $\mathbb{F}^n$ can be decomposed as a product of reflections

$$M(X): Y \mapsto Y - [2(X, Y)/(X, X)]X \tag{4}$$

with $(X, X) \neq 0$. It can be shown[3,4] that the spinor norm $\mathrm{Spin}: O(p, q) \to \mathbb{F}_*/\mathbb{F}_*^{(2)}$ is given by

$$\mathrm{Spin}(M) \equiv \prod_{i=1}^{s} (X_i, X_i) \pmod{\mathbb{F}_*^{(2)}} \tag{5}$$

for each pseudo-orthogonal matrix M over $\mathbb{F}$ which has been decomposed as a product of reflections $M = M(X_1) \ldots M(X_s)$. Here $\mathbb{F}_*$ denotes the multiplicative group of $\mathbb{F}$, and $\mathbb{F}_*^{(2)}$ denotes its subgroup of square elements. For the real field $\mathbb{R}$, each positive number is a square and therefore sign: $\mathbb{R}_* \to Z_2$ is a homomorphism with $\mathbb{R}_*^{(2)}$ as kernel. An equivalent definition of the spinor norm is thus given by

$$\mathrm{spin}(M) = \prod_{i=1}^{s} \mathrm{sign}((X_i, X_i)). \tag{6}$$

Such a formula does not provide a simple method for calculating the spinor norm since it requires the decomposition of M into reflections. Some time ago, Zassenhaus[3] proved that the spinor norm is equal to $\det(I_n + M) \pmod{\mathbb{F}_*^{(2)}}$ [i.e., $\mathrm{sign}(\det(I_n + M))$ in the real case] if $\det(I_n + M) \neq 0$. If, however, this determinant equals zero, further computations are needed.

It is well known in the physical literature that, for $O(3, 1)$, the spinor norm is simply given by the sign of D ($p = 3$, $q = 1$) in the block form (1) of M. It is the aim of this note to show that this very simple method can be generalized to all other real pseudo-orthogonal groups. More precisely we prove the following theorem.

Theorem: The spinor norm $\mathrm{spin}: O(p, q) \to \mathbb{Z}_2$ is given by

$$\mathrm{spin}(M) = \mathrm{sign}(\det D) = \mathrm{sign}(\det A) \times \det M \tag{7}$$

[where A and D are defined by (1)].

Part 1: $|\det A| \geq 1$, $|\det D| \geq 1$ (8)

Proof: By virtue of Eqs. (2), (3) we have

$$\det(A^T A) = (\det A)^2 = \det(I_p + C^T C),$$

$$\det(D^T D) = (\det D)^2 = \det(I_q + B^T B);$$

but $C^T C$ and $B^T B$ are two symmetric nonnegative definite matrices, thus $\det(I_p + C^T C) \geq 1$ and $\det(I_q + B^T B) \geq 1$.

 0022-2488/78/1906-1265$1.00

Since the mappings $M \mapsto \det A$ and $M \mapsto \det D$ are polynomial and therefore continuous, it follows from the inequalities (8) that the sign of $\det A$ and of $\det D$ are constant for all matrices belonging to a connected component of $O(p,q)$. We have furthermore:

Part 2: The mappings $O(p,q) \to Z_2$ defined by

$$\varphi_1: M \mapsto \operatorname{sign}(\det A), \quad \varphi_2: M \mapsto \operatorname{sign}(\det D)$$

are group homomorphisms.

Proof: We give the proof only for the mapping φ_2, the one for φ_1 being exactly the same. Since $I_n \mapsto 1 \in Z_2$ is evident, it remains to prove that $\varphi_2(M_1)\varphi_2(M_2) = \varphi_2(M_1M_2)$. Considering the block forms (1) for M_1 and M_2 and using the matrix multiplication for blocks, we obtain

$$\varphi_2(M_1M_2) = \operatorname{sign}(\det(D_1D_2 + C_1B_2)).$$

Since by virtue of (8) the matrices D_1 and D_2 are invertible, we have

$$\det(D_1D_2 + C_1B_2) = \det D_1 \det D_2 \det(I_q + D_1^{-1}C_1B_2D_2^{-1}),$$

and it remains to show that $\det(I_q + D_1^{-1}C_1B_2D_2^{-1}) \geq 0$ [and therefore > 0 by virtue of (8)]. This inequality is true if the matrix $H = D_1^{-1}C_1B_2D_2^{-1}$ is bounded with bound $\|H\| \leq 1$. Indeed, if H is such a matrix, its spectrum is entirely contained in the unit disk and if the spectrum contains real eigenvalues, these are contained in the interval $[-1,1]$. Therefore, $X_H(-1) = \det(I_q + H) \geq 0$ (here $X_H(t) = \det(-tI_q + H)$ denotes the characteristic polynomial of H). Actually the stronger property

$$(x, Hy)_2 \leq [(x,x)_2(y,y)_2]^{1/2}, \quad x,y \in \mathbb{R}^q \tag{9}$$

holds, from which follows the boundedness of H. Indeed, by using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}(x, Hy)_2 &= (x, D_1^{-1}C_1B_2D_2^{-1}y)_2 = (C_1^TD_1^{T-1}x, B_2D_2^{-1}y)_1 \\ &\leq [(C_1^TD_1^{T-1}x, C_1^TD_1^{T-1}x)_1(B_2D_2^{-1}y, B_2D_2^{-1}y)_1]^{1/2}.\end{aligned}$$

Then, from Eqs. (2), (3), we obtain respectively

$$\begin{aligned}0 &\leq (C_1^TD_2^{T-1}x, C_1^TD_2^{T-1}x)_1 \\ &= (x, D_1^{-1}C_1C_1^TD_1^{T-1}x)_2 = (x, D_1^{-1}(D_1D_1^T - I_q)D_1^{T-1}x)_2 \\ &= (x,x)_2 - (D_1^{T-1}x, D_1^{T-1}x)_2 \leq (x,x)_2,\end{aligned}$$

$$\begin{aligned}0 &\leq (B_2D_2^{-1}y, B_2D_2^{-1}y)_1 \\ &= (y, D_2^{T-1}B_2^TB_2D_2^{-1}y)_2 = (y, D_2^{T-1}(D_2^TD_2 - I_q)D_2^{-1}y)_2 \\ &= (y,y)_2 - (D_2^{-1}y, D_2^{-1}y)_2 \leq (y,y)_2,\end{aligned}$$

which proves the property (9) and consequently the homomorphism property of the mapping φ_2.

Part 3: $\operatorname{spin}(M) = \varphi_2(M) = \varphi_1(M)\det M$.

Proof: Since φ_1 and φ_2 are group homomorphisms, the only thing to show, by virtue of (6), is that

$$\operatorname{spin}(M(X)) = \varphi_2(M(X)) = \varphi_1(M(X))\det M(X),$$

where $M(X)$ is any reflection. That is, in block form

$$M(X) = \begin{pmatrix} I_p - 2x_1x_1^T/(X,X) & 2x_1x_2^T/(X,X) \\ -2x_2x_1^T/(X,X) & I_q + 2x_2x_2^T/(X,X) \end{pmatrix}.$$

Now, on the one hand

$$\begin{aligned}&\operatorname{sign}\left(\det\left(I_q + \frac{2x_2x_2^T}{(X,X)}\right)\right) \\ &= \operatorname{sign}\left(1 + \frac{2(x_2,x_2)_2}{(X,X)}\right) = \operatorname{sign}\left(\frac{(x_1,x_1)_1 + (x_2,x_2)_2}{(X,X)}\right) \\ &= \operatorname{sign}((X,X)) = \operatorname{spin}(M(X)),\end{aligned}$$

which proves the first equality. On the other hand,

$$\begin{aligned}&\operatorname{sign}\left(\det\left(I_p - \frac{2x_1x_1^T}{(X,X)}\right)\right) \\ &= \operatorname{sign}\left(1 - \frac{2(x_1,x_1)_1}{(X,X)}\right) = \operatorname{sign}\left(-\frac{(x_1,x_1)_1 + (x_2,x_2)_2}{(X,X)}\right) \\ &= -\operatorname{sign}((X,X)) = -\operatorname{spin}(M(X)).\end{aligned}$$

However, since $\det: O(p,q) \to Z_2$ is a group homomorphism and $\det M(X) = -1$, the second equality is proved.

REMARKS

1. The main result (7) and the inequalities (8) can be trivially extended to the pseudo-orthogonal groups $O(G, \mathbb{R})$ associated with the symmetric matrices of the form

$$G = \begin{pmatrix} G_1 & 0 \\ 0 & -G_2 \end{pmatrix},$$

where G_1 and G_2 are two symmetric positive definite matrices of degree p and q respectively ($pq \neq 0$).

2. These results can also be extended to all pseudo-orthogonal groups $O(G, \mathbb{F})$ over a field $\mathbb{F}$ for which the definition (5) of the spinor norm can be replaced by the definition (6). This is possible for all complete ordered fields. This, for example, is not the case for the field of rational numbers since not every positive rational number is a square.

ACKNOWLEDGMENT

I wish to thank Professor H. Zassenhaus for his valuable comments.

[1] R. Lipschitz, *Untersuchung über die Summen von Quadraten* (Bonn, 1886).

[2] N. Bourbaki, *Elêments de mathêmatiques, Algêbre* (Hermann, Paris, 1959), Livre 2, Chap. 9.

[3] H. Zassenhaus, Arch. Math. 13, 434 (1962).

[4] J. A. Dieudonnê, *La gêomêtrie des groupes classiques* (Springer, Ergebnisse Bd 5, 1975).

Boundary condition solutions of the generalized Feller equation

Siegfried H. Lehnigk

MIRADCOM, Redstone Arsenal, Alabama 35809
and Department of Mathematics, University of Alabama in Huntsville, Huntsville, Alabama 35807

Solutions of the generalized one-dimensional autonomous parabolic Feller equation for given boundary conditions are established. Since a basic solution is known, the Green–Riemann technique is used. It leads to two Green–Riemann limit functions relative to subsets of the space of two parameters connected with the equation. Properties of these limit functions are discussed. These functions are then used to establish boundary condition solutions in the form of unilateral convolutions involving as boundary conditions those functions which are summable over every nonnegative compact interval. Finally, it is shown that, relative to two subsets of the space of two equation parameters, there exist initial and boundary condition solutions.

1. INTRODUCTION

Initial condition solutions of the generalized Feller equation (of Fokker–Planck type)

$$l(z)=A(x)z_{xx}+B(x)z_x+C(x)z-z_t=0,\quad z=z(x,t), \tag{1.1a}$$

on the domain $x>0$, $t>0$, with coefficients

$$\begin{aligned}&A(x)=\alpha x^{\lambda+1},\quad \alpha>0,\quad \lambda<1,\\ &B(x)=\beta_1 x^{\lambda}+\beta_2 x,\quad \beta_{1,2}\in\mathbb{R},\\ &C(x)=\rho x^{\lambda-1}+\beta_2,\quad \rho=\lambda[\beta_1-\alpha(1+\lambda)],\end{aligned} \tag{1.1b}$$

have been established in Ref. 1. They are singular integrals of the form

$$z_*(x,t)=\int_0^\infty v^*(x,t;y)f(y)\,dy,\quad x>0,\quad 0<t<t_0, \tag{1.2}$$

under the assumption that $f(y)\in\mathcal{L}$ $(y_0\le y\le y_1)$ for every y_0, y_1, $0\le y_0<y_1<\infty$ (Lebesgue integrable over every compact nonnegative interval), and $v^*(x_0,t_0;y)f(y)\in\mathcal{L}$ $(0\le y<\infty)$ for some $x_0>0$, $t_0>0$. The kernel v^* in (1.2) is given by

$$\begin{aligned}v^*(x,t;y)=(1-\lambda)x^{-1}\varphi_0^{(1-\lambda)(1-\nu/2)}\eta_0^{\nu(1-\lambda)/2}\bar{\mathcal{I}}_\nu(r_0)\\ \times\exp(-\varphi_0^{1-\lambda}-\eta_0^{1-\lambda}),\end{aligned} \tag{1.3}$$

for $x>0$, $y>0$, $t>0$, where

$$\begin{aligned}&\nu=(1-\lambda)^{-1}(\alpha^{-1}\beta_1-1-2\lambda),\quad r_0=2(\varphi_0\eta_0)^{(1-\lambda)/2},\\ &\varphi_0=\xi_0\exp\beta_2 t,\quad \xi_0=xb_0^{-1},\quad \eta_0=yb_0^{-1},\end{aligned} \tag{1.4}$$

$$b_0=b_0(t)=\begin{cases}[\alpha\beta_2^{-1}(1-\lambda)(-1+\exp[(1-\lambda)\beta_2 t])]^{(1-\lambda)^{-1}}, & \beta_2\neq 0,\\ [\alpha(1-\lambda)^2 t]^{(1-\lambda)^{-1}}, & \beta_2=0.\end{cases} \tag{1.5}$$

Furthermore, in (1.3),

$$\bar{\mathcal{I}}_\nu(r_0)=\begin{cases}I_\nu(r_0) & \text{if } \nu>0,\\ I_{-\nu}(r_0) & \text{if } \nu<1,\end{cases} \tag{1.6}$$

where I_μ is the modified Bessel function of the first kind of order μ.

The kernel $v^*(x,t;y)$ of (1.2) has been obtained from the basic solution

$$\begin{aligned}v(x,t;y,s)=x^{-1}b^{1+\lambda}\varphi^{(1-\lambda)(1-\nu/2)}\eta^{1+\lambda+\nu(1-\lambda)/2}\\ \times\bar{\mathcal{I}}_\nu(r)\exp(-\varphi^{1-\lambda}-\eta^{1-\lambda}),\end{aligned} \tag{1.7}$$

of $l(z)=0$, $x>0$, $y>0$, $t>s\ge 0$, in which

$$\begin{aligned}&r=2(\varphi\eta)^{(1-\lambda)/2},\quad \varphi=\xi\exp\beta_2(t-s),\\ &\xi=xb^{-1},\quad \eta=yb^{-1},\quad b=b(t,s)=b_0(t-s).\end{aligned}$$

The assertion that $v(x,t;y,s)$ is a basic solution of $l(z)=0$ means that, as a function of x and t, $v(x,t;y,s)$ is a solution of $l(z)=0$ and, as a function of y and s, it is a solution of the adjoint equation.

The functions v^* and v are related by

$$v^*(x,t;y)=(1-\lambda)y^{-1-\lambda}v(x,t;y,0).$$

For $\nu<1$ [in which case $\bar{\mathcal{I}}_\nu=I_\nu$ according to (1.6)], $v^*(x,t;y)$ has the property that

$$v^*(x,t;0+)=z_0(x,t)=\frac{1-\lambda}{\Gamma(1-\nu)}x^{-1}\varphi_0^{(1-\lambda)(1-\nu)}\exp(-\varphi_0^{1-\lambda}). \tag{1.8}$$

Certain aspects and consequences of (1.2) and (1.8) have been discussed in Refs. 2 and 3.

The objective of this paper is to present solutions of the generalized Feller equation (1.1) with prescribed behavior (boundary condition) along the positive t axis. The results together with those of Ref. 1 will lead to Cauchy solutions of (1.1), i.e., solutions of (1.1) for given initial and boundary conditions, in subsets of the space of the parameters $\lambda<1$, , $\nu\in\mathbb{R}$.

The boundary condition solutions will be given as unilateral convolutions

$$\begin{aligned}&z_p(x,t)=p(x,t)\overset{t}{\underset{0}{*}}g(t),\quad x>0,\ t>0,\ \lambda<0,\ \nu=-\lambda(1-\lambda)^{-1},\\ &z_q(x,t)=q(x,t)\overset{t}{\underset{0}{*}}g(t),\quad x>0,\ t>0,\ \lambda=0,\ \nu<0,\end{aligned}$$

for functions $g(t)$ of a suitable function class.

2. GREEN–RIEMANN LIMITS

By a boundary condition solution of $l(z)=0$ we mean

 0022-2488/78/1906-1267$1.00

a function $w(x,t)$ defined on $x>0$, $t>0$, such that $l(w)\equiv 0$ and $w(0+,t)=g(t)$ a.e., $t>0$, for given $g(t)$ of a suitable function space.

Since a basic solution $v(x,t;y,s)$ [(1.7)] of $l(z)=0$ is known, the existence problem of boundary condition solutions is most easily treated by means of the Green–Riemann method. To use that method, it is advantageous to consider, instead of $l(z)$, the operator

$$L(z)=A^{-1}(y)l(z)$$
$$=z_{yy}+B^*(y)z_y+C^*(y)z-D^*(y)z_s,\quad z=z(y,s).$$

The adjoint of $L(z)$ is

$$M(u)=u_{yy}-B^*(y)u_y+\left[C^*(y)-\frac{d}{dy}B^*(y)\right]u+D^*(y)u_s,$$
$$u=u(y,s).$$

Let $\mathcal{R}$ be a rectangular domain in the Cartesian (y,s) plane with sides parallel to the y and s axes and lower left-hand corner at the origin. Furthermore, let z be a solution of $L(z)=0$ and let u be a solution of the adjoint equation $M(u)=0$. Then

$$\int\int_{\mathcal{R}}[uL(z)-zM(u)]\,dy\,ds=\int_{\text{sides of }\mathcal{R}}[P\,ds-Q\,dy]=0,$$

where

$$P=u\frac{\partial z}{\partial y}-\left[\frac{\partial u}{\partial y}-B^*(y)u\right]z,$$
$$Q=-A^{-1}(y)\,zu.$$

As to the existence of boundary condition solutions only the behavior of $\lim P$ as $y\downarrow 0$ is of interest. If we take for u the basic solution $v(x,t;y,s)$ of $L(z)=0$ given by (1.7) and if we expect to obtain boundary condition solutions at least for identically constant limit functions $g(t)$ (as $x\downarrow 0$), it follows that boundary condition solutions exist if and only if there exists a function $\tilde{p}(x,t)$ from $x>0$, $t>0$, into $\mathbb{R}$ such that

$$\lim\int_0^t\tilde{p}(x,\tau)\,d\tau=c\quad\text{as } x\downarrow 0,\tag{2.1}$$

where c is a finite nonzero constant, and such that

$$\lim\left[\frac{\partial v}{\partial y}-B^*(y)v\right]=\tilde{p}(x,t-s),\quad t>s\geq 0,\quad\text{as } y\downarrow 0.\tag{2.2}$$

For the given Eq. (1.1) we have

$$\frac{\partial v}{\partial y}-B^*(y)v=\begin{cases}b^{-1}\eta^{-1}v[-\lambda+\frac{1}{2}(1-\lambda)rI_\nu^{-1}(r)I_{\nu+1}(r)-(1-\lambda)\eta^{1-\lambda}-\alpha^{-1}\beta_2 b^{1-\lambda}\eta^{1-\lambda}], & Z_\nu=I_\nu,\\ b^{-1}\eta^{-1}v[-\lambda-\nu(1-\lambda)+\frac{1}{2}(1-\lambda)rI_{-\nu}^{-1}(r)I_{-\nu+1}(r)-(1-\lambda)\eta^{1-\lambda}-\alpha^{-1}\beta_2 b^{1-\lambda}\eta^{1-\lambda}], & Z_\nu=I_{-\nu}.\end{cases}$$

This shows that the function $(\partial v/\partial y)-B^*(y)v$ goes to 0 or $+\infty$ as $y\downarrow 0$ except in the following cases.

(1) $\lambda<0$, $\lambda+\nu(1-\lambda)=0$, $Z_\nu=I_\nu$ in (1.7):

$$\frac{\partial v}{\partial y}-B^*v-\tilde{p}_1(x,t-s)=-\frac{\lambda}{\Gamma(1+\nu)}x^{-1}b^\lambda\varphi^{1-\lambda}\exp(-\varphi^{1-\lambda}).\tag{2.3}$$

(2) $\lambda=0$, $\nu<0$, $Z_\nu=I_{-\nu}$ in (1.7):

$$\frac{\partial v}{\partial y}-B^*v-\tilde{p}_2(x,t-s)=\frac{1}{\Gamma(-\nu)}x^{-1}\varphi^{1-\nu}e^{-\varphi}.\tag{2.4}$$

(3) $\lambda=0$, $0<\nu<1$, $Z_\nu=I_{-\nu}$ in (1.7):

$$\frac{\partial v}{\partial y}-B^*v-\tilde{p}_3(x,t-s)=\frac{1}{\Gamma(-\nu)}x^{-1}\varphi^{1-\nu}e^{-\varphi}.$$

This last case, however, is of no further interest here since the condition (2.1) is not satisfied for $\tilde{p}_3(x,t)$.

It is easily verified directly that the functions $\tilde{p}_1(x,t)$ and $\tilde{p}_2(x,t)$ are solutions of $l(z)=0$ on $x>0$, $t>0$. As a matter of fact, they are the only solutions of $l(z)=0$ of the form

$$x^\mu b_0^\sigma\varphi_0^\tau\exp(-\varphi_0^{1-\lambda})\tag{2.5}$$

for which (2.1) holds. Substitution of (2.5) into (1.1) leads to $\mu=-1$ and to a quadratic equation for τ which has the roots $\tau_1=1-\lambda$, $\tau_2=1-\lambda-\nu(1-\lambda)$. The corresponding σ values are $\sigma_1=\lambda$, $\sigma_2=0$. For (2.1) to hold, it is necessary that the exponents in (2.5) satisfy the condition $\mu+\sigma+1-\lambda=0$ which is trivial for $\sigma=\sigma_1=\lambda$ and which leads to $\lambda=0$ for $\sigma=\sigma_2=0$. Thus, $\mu=-1$, $\sigma_1=\lambda$, $\tau_1=1-\lambda$, leads to (2.3) and $\mu=-1$, $\sigma_2=0$, $\tau_2=1-\nu$, leads to (2.4).

3. THE FUNCTION $p(x,t)$

Using the function $\tilde{p}_1(x,t)$ given in (2.3), we define on $x>0$, $-\infty<t<\infty$, the nonnegative function

$$p(x,t)=\begin{cases}-\alpha\nu\lambda^{-1}(1-\lambda)^2\tilde{p}_1(x,t)\\ \quad=[\alpha(1-\lambda)^2/\Gamma(\nu)]x^{-1}b_0^\lambda\varphi_0^{1-\lambda}\exp(-\varphi_0^{1-\lambda}), & t>0,\\ 0, & t\leq 0,\end{cases}\tag{3.1}$$

where $\lambda<0$, $\nu=-\lambda(1-\lambda)^{-1}$. The definitions of the variables are given in (1.4) and (1.5).

The function $p(x,t)$ may also be expressed in terms of the function $z_0(x,t)$ given in (1.8). For $t>0$, this leads to

$$p(x,t)=\alpha(1-\lambda)\frac{\Gamma((1-\lambda)^{-1})}{\Gamma(\nu)}b_0^\lambda\varphi_0^{-\lambda}z_0(x,t),\quad t>0.$$

In particular, for $\lambda=-1$, $\nu=\frac{1}{2}$, $\beta_1=\beta_2=0$, and $t>0$, we obtain

$$p_H(x,t)=\frac{4\alpha}{\pi^{1/2}}\frac{x}{(4\alpha t)^{3/2}}\exp[-x^2(4\alpha t)^{-1}],\quad t>0,$$

which is the well-known singular solution of the heat equation $\alpha z_{xx}-z_t=0$.

We prove two basic theorems about the function $p(x,t)$.

Theorem 3.1: $p(x,t)\in C^\infty$, $x>0$, $-\infty<t<\infty$.

Proof: For $t\neq 0$, the statement follows directly from the explicit form of $p(x,t)$ in (3.1).

For $t<0$, all derivatives of $p(x,t)$ go to 0 as $(x,t)\to(x_0,0-)$ for $x_0>0$.

Let $t>0$. It is convenient to express $p(x,t)$ in the form

$$p(x,t)=x^{-\lambda}[y-\alpha^{-1}(1-\lambda)^{-1}\beta_2]^{\nu}\,y\exp(-x^{1-\lambda}y),$$

with

$$y=y(t)=b_0^{-1+\lambda}(t)\exp(1-\lambda)\beta_2 t=b_0^{-1+\lambda}(t)+\alpha^{-1}(1-\lambda)^{-1}\beta_2,$$

and

$$\frac{dy}{dt}=-\alpha(1-\lambda)^2y^2+\beta_2(1-\lambda)y.$$

The constant factor of p has been disregarded. We see now that any differentiation of $p(x,t)$ produces an expression of the same structure as $p(x,t)$ itself. In particular, the exponential factor $\exp(-x^{1-\lambda}y)$ will always occur. Therefore, since $y\uparrow\infty$ as $t\downarrow 0$, each derivative of $p(x,t)$ goes to 0 as $(x,t)\to(x_0,0+)$, $x_0>0$. This completes the proof.

Obviously, the function $p(x,t)$ is a solution of $l(z)=0$, $\lambda<0$, $\nu=-\lambda(1-\lambda)^{-1}$ for $x>0$, $t<0$, and it has been shown at the end of Sec. 2 that it is also a solution of $l(z)=0$ for $x>0$, $t>0$. We prove now

Theorem 3.2: $p(x,t)$ is a solution of $l(z)=0$ for $x>0$, $-\infty<t<\infty$.

Proof: Define the function

$$P(x,t)=A(x)\frac{\partial^2 p}{\partial x^2}+B(x)\frac{\partial p}{\partial x}+C(x)p$$

with A, B, and C given in (1.1b). Then

$$l(p)=P(x,t)-\frac{\partial p(x,t)}{\partial t}\equiv 0 \tag{3.2}$$

for $x>0$ and $t\neq 0$. Since $\lim[P(x,t)-\partial p(x,t)/\partial t]=0$ as $(x,t)\to(x_0,0-)$, $x_0>0$, we have to show that this limit relation also holds as $(x,t)\to(x_0,0+)$. Because of (3.2) it suffices to show that $P(x,t)\to 0$ as $(x,t)\to(x_0,0+)$, $x_0>0$.

Let $t>0$. Introducing the new variable

$$y=b_0^{-1+\lambda}(t)\exp(1-\lambda)\beta_2 t,\quad y\uparrow\infty \text{ as } t\downarrow 0,$$

and disregarding a constant factor, we can write $P(x,t)$ as

$$\begin{aligned}P(x,t)=x^{-\lambda}&[(y-\alpha^{-1}(1-\lambda)^{-1}\beta_2)y^{-1}]^{\nu}y^{\mu}\exp(-x^{1-\lambda}y)\\&\times[(3\alpha\lambda-\beta_1)(1-\lambda)y^{-1}+\alpha(1-\lambda)^2x^{1-\lambda}\\&-\beta_2(1-\lambda)x^{1-\lambda}y^{-1}+\beta_2(1-\lambda)y^{-2}]y^{-1}\end{aligned}$$

with $\mu=(4-5\lambda)(1-\lambda)^{-1}>0$. There exists a positive constant c which dominates the first bracketed factor. Furthermore, we observe the inequality

$$0\le y^{\mu}\exp(-ry)\le(\mu e^{-1}r^{-1})^{\mu},\quad 0\le y<\infty,\ \mu>0,\ r>0.$$

If we apply it with $\mu=(4-5\lambda)(1-\lambda)^{-1}>0$, $r=x^{1-\lambda}>0$, and if we restrict x and t such that $|x-x_0|<\delta$, $0<t<\delta$, and $0<\delta<\frac{1}{2}x_0$, then

$$\begin{aligned}|P(x,t)|<c(\mu e^{-1})^{\mu}(2x_0^{-1})^{4(1-\lambda)}&[|3\alpha\lambda-\beta_1|(1-\lambda)y^{-1}\\&+\alpha(1-\lambda)^2(\tfrac{3}{2}x_0)^{1-\lambda}+|\beta_2|(1-\lambda)(\tfrac{3}{2}x_0)^{1-\lambda}y^{-1}\\&+|\beta_2|(1-\lambda)y^{-2}]y^{-1}\end{aligned} \tag{3.3}$$

if y is sufficiently large.

Given now $\epsilon>0$, we shall have $|P(x,t)|<\epsilon$ for $|x-x_0|<\delta$, $0<t<\delta$, if $\delta>0$ is sufficiently small. This follows from the fact that $[\cdots]y^{-1}$ in (3.3) goes to 0 as $y^{-1}\downarrow 0$, i.e., as $t\downarrow 0$. Thus, the proof is complete.

It is useful to list several properties of the function $p(x,t)$ some of which are easily verified by inspection or implied by earlier results.

(1) $p(x,t)\downarrow 0$ as $t\downarrow 0$, $x>0$.

(2) $p(x,t)\downarrow 0$ as $x\downarrow 0$, $t>0$. In other words, $p(x,t)$ is a singular solution of $l(z)=0$ (in the sense of Doetsch[4]).

(3) $p(x,t)\downarrow 0$ as $t\uparrow\infty$, $x>0$.

(4) $x>0\Rightarrow p(x,t)$ has exactly one maximum in $0<t<\infty$ which is located at

$$t_m=\begin{cases}(1-\lambda)^{-1}\beta_2^{-1}\log\{\frac{1}{2}\lambda^{-1}(1-\alpha^{-1}\beta_2x^{1-\lambda})\\\quad+[\frac{1}{4}\lambda^{-2}(1-\alpha^{-1}\beta_2x^{1-\lambda})^2+\nu^{-1}]^{1/2}\},&\beta_2\neq 0,\\ \alpha^{-1}(1-2\lambda)^{-1}(1-\lambda)^{-1}x^{1-\lambda},&\beta_2=0,\end{cases}$$

and $t_m\downarrow 0$ as $x\downarrow 0$. Thus, for $x>0$, $p(x,t)\in\uparrow$, $0\le t\le t_m$, and $p(x,t)\in\downarrow$, $t_m\le t<\infty$ ($\uparrow$ = set of nondecreasing functions). Furthermore, since $\varphi_0(t_m)\to[(1-2\lambda)(1-\lambda)^{-1}]^{(1-\lambda)^{-1}}$ as $x\downarrow 0$ and since $b_0(t_m)\downarrow 0$ as $x\downarrow 0$, it follows that $p(x,t_m)\uparrow\infty$ as $x\downarrow 0$.

(5) There exists a constant $k_p>0$ such that

$$0<t_m p(x,t_m)\le k_p,\quad 0<x<\infty. \tag{3.4}$$

This can be seen as follows. For $\beta_2\neq 0$, let $y=y(x)=\exp(1-\lambda)\beta_2 t_m$. Then

$$\begin{aligned}t_m p(x,t_m)=\frac{(1-\lambda)^{-\nu}}{\Gamma(\nu)}&y^{(1-\lambda)^{-1}}(y-1)(\log y)(1-\lambda-\lambda y)^{\nu}\\&\times\exp[-(1+\nu y)].\end{aligned}$$

$0<x<\infty$ implies $0<y<1$ if $\beta_2<0$ and $1<y<\infty$ if $\beta_2>0$. The function $t_m p(x,t_m)\in C$ $(0<y<\infty)$ if we define its value at $y=1$ by

$$\frac{(1+\nu)^{\lambda}}{\Gamma(\nu)}\exp[-(1+\nu)]=\lim t_m p(x,t_m)\ \text{ as } y\to 1.$$

Furthermore, $t_m p(x,t_m)\downarrow 0$ as $y\downarrow 0$ and as $y\uparrow\infty$. Consequently, $t_m p(x,t_m)$ is bounded.

If $\beta_2=0$, we obtain

$$t_m p(x,t_m)\equiv\frac{(1+\nu)^{\nu}}{\Gamma(\nu)}\exp[-(1+\nu)].$$

(6) $x>0\Rightarrow$

$$\int_0^{\infty}p(x,t)\,dt=\begin{cases}\exp[-\alpha^{-1}\beta_2(1-\lambda)^{-1}x^{1-\lambda}],&\beta_2>0,\\ [1/\Gamma(\nu)]\,\Gamma(\nu;-\alpha^{-1}\beta_2(1-\lambda)^{-1}x^{1-\lambda})\\\quad\times\exp[-\alpha^{-1}\beta_2(1-\lambda)^{-1}x^{1-\lambda}],&\beta_2<0,\\ 1,&\beta_2=0,\end{cases}$$

where

$$\Gamma(\nu;a)=\int_a^{\infty}\sigma^{\nu-1}e^{-\sigma}\,d\sigma,\quad a\ge 0.$$

This can be verified by reduction of $\int_0^{\infty}p(x,t)$ to the gamma function integral by means of the substitution $(xb_0^{-1})^{1-\lambda}=\sigma$.

(7) $x > 0 \Rightarrow$

$$P(x) = \int_0^\infty p(x,t)\,dt \in \downarrow, \quad \lim P(x) = 1 \text{ as } x \downarrow 0. \quad (3.5)$$

The limit relation is evident from (3.4). That $P(x)$ is nonincreasing is obvious from (3.4) for $\beta_2 \geq 0$. For $\beta_2 < 0$, set $-\alpha^{-1}\beta_2(1-\lambda)^{-1}x^{1-\lambda} = y$. Then, from (3.4),

$$P(x) = e^y[\Gamma(\nu) - \int_0^y \sigma^{\nu-1}e^{-\sigma}\,d\sigma].$$

Differentiating with respect to x and disregarding a positive constant factor, we see that dP/dx is negative for $x > 0$ since

$$\Gamma(\nu) < \int_0^y \sigma^{\nu-1}e^{-\sigma}\,d\sigma + y^{\nu-1}e^y, \quad y > 0.$$

The two properties (3.5) of $p(x,t)$ may be combined into

$$0 < \int_0^\infty p(x,t)\,dt \leq 1, \quad 0 < x < \infty.$$

(8) $x > 0$, $t_0 > 0 \Rightarrow$

$$\lim \int_0^{t_0} p(x,t)\,dt = 1 \text{ as } x \downarrow 0.$$

This limit relation can be verified by transformation of the original integral into the incomplete gamma function integral by means of the substitution used under 6.

(9) There exists a constant $k > 0$ such that

$$\int_0^\infty t\left|\frac{\partial p(x,t)}{\partial t}\right| dt < k, \quad 0 < x < \infty.$$

To see this, we observe that

$$\frac{\partial p(x,t)}{\partial t}\begin{cases} \geq 0, & 0 \leq t \leq t_m, \\ \leq 0, & t_m \leq t < \infty. \end{cases}$$

Therefore,

$$\int_0^\infty t\left|\frac{\partial p(x,t)}{\partial t}\right| dt = \int_0^{t_m} t\frac{\partial p(x,t)}{\partial t}\,dt - \int_{t_m}^\infty t\frac{\partial p(x,t)}{\partial t}\,dt,$$

and partial integration leads to

$$\int_0^\infty t\left|\frac{\partial p(x,t)}{\partial t}\right| dt = 2t_m p(x,t_m) - \int_0^{t_m} p(x,t)\,dt + \int_{t_m}^\infty p(x,t)\,dt.$$

Consequently,

$$\int_0^\infty t\left|\frac{\partial p(x,t)}{\partial t}\right| dt < 2t_m p(x,t_m) + \int_0^\infty p(x,t)\,dt < 2k_p + 1 = k, \quad 0 < x < \infty.$$

4. BOUNDARY CONDITION SOLUTIONS BY THE p-CONVOLUTION

We introduce the unilateral convolution

$$z_p(x,t) = p(x,t) * g(t) = \int_0^t p(x,t-s)g(s)\,ds, \quad x > 0, \ t > 0, \quad (4.1)$$

where the convolution kernel $p(x,t)$ is given in (3.1). Admissible functions $g(t)$ are those wich are summable in every finite nonnegative interval.

Theorem 4.1: $x > 0$, $t > 0$,

$z_p(x,t) = p(x,t) * g(t)$, $g(t) \in \mathcal{L}(t_0 \leq t \leq t_1)$,

$0 \leq t_0 < t_1 < \infty$,

$\Rightarrow z_p(x,t)$ exists and $z_p(x,0+) = 0$.

Proof: Because of the continuity of $p(x,t)$ and the assumption about $g(t)$, the existence of $z_p(x,t)$ is obvious. Next, since there exists a positive constant $p_0 = p_0(x)$ such that $0 \leq p(x,\sigma) \leq p_0$ for $x > 0$ and $-\infty < \sigma < \infty$, we have the inequality

$$|z_p(x,t)| = |p(x,t) * g(t)| \leq p_0 \int_0^t |g(s)|\,ds, \quad x > 0,$$

which leads directly to the limit relation of the theorem.

The next theorem deals with the inversion of the transformation (4.1).

Theorem 4.2: $x > 0$, $t > 0$,

$z_p(x,t) = p(x,t) * g(t)$, $g(t) \in \mathcal{L}(t_0 \leq t \leq t_1)$,

$0 \leq t_0 < t_1 < \infty \quad \Rightarrow z_p(0+,t) = g(t)$ a.e.

Proof: Because of the summability assumption on g, almost every point $t \geq 0$ is a Lebesgue point of g. Therefore, to prove the theorem, it is sufficient to prove its limit relation for a positive Lebesgue point t of g. This will be done in analogy to the proof given in Ref. 5, p. 78, for the special case of the heat equation ($\lambda = -1$, $\nu = \frac{1}{2}$).

We rewrite (4.1) as

$$z_p(x,t) = \int_0^t p(x,\sigma)g(t-\sigma)\,d\sigma.$$

Since

$$\lim \int_0^t p(x,\sigma)\,d\sigma = 1 \text{ as } x \downarrow 0, \ t > 0,$$

the limit relation of the theorem is equivalent to

$$\lim[z_p(x,t) - g(t)] = \lim \int_0^t p(x,\sigma)[g(t-\sigma) - g(t)]\,d\sigma \text{ as } x \downarrow 0.$$

Let

$$a(\sigma) = \int_0^\sigma [g(t-\tau) - g(t)]\,dt. \quad (4.2)$$

Then

$$z_p(x,t) - g(t) = \int_0^t p(x,\sigma)\,da(\sigma),$$

which, by partial integration and since $p(x,0+) = 0$, changes into

$$z_p(x,t) - g(t) = p(x,t)a(t) - \int_0^t \frac{\partial p(x,\sigma)}{\partial \sigma}a(\sigma)\,d\sigma. \quad (4.3)$$

Since the first term at the right-hand side goes to 0 as $x \downarrow 0$, it remains to show that the integral does the same.

Let $\epsilon > 0$ be given. Since t, by assumption, is a positive Lebesgue point of g, there exists δ, $0 < \delta < t$, such

that

$$\frac{1}{\sigma}\int_{t-\sigma}^{t}|g(\tau)-g(t)|\,d\tau<\epsilon,\quad 0<\sigma\le\delta.$$

A change of variable in (4.2) shows that

$$\sigma^{-1}a(\sigma)=\frac{1}{\sigma}\int_{t-\sigma}^{t}[g(\tau)-g(t)]\,d\tau,\quad \sigma>0.$$

Therefore,

$$\sigma^{-1}|a(\sigma)|<\epsilon,\quad 0<\sigma\le\delta.$$

We now write the integral in (4.3) as

$$\int_0^t\frac{\partial p(x,\sigma)}{\partial\sigma}a(\sigma)\,d\sigma=\int_0^{\delta}+\int_{\delta}^{t}. \tag{4.4}$$

The first integral at the right-hand side can be estimated as follows:

$$\left|\int_0^{\delta}\frac{\partial p(x,\sigma)}{\partial\sigma}a(\sigma)\,d\sigma\right|$$
$$\le\int_0^{\delta}\sigma\left|\frac{\partial p(x,\sigma)}{\partial\sigma}\right|\sigma^{-1}a(\sigma)\,d\sigma<\epsilon\int_0^{\delta}\sigma\left|\frac{\partial p(x,\sigma)}{\partial\sigma}\right|d\sigma$$
$$<\epsilon\int_0^{\infty}\sigma\left|\frac{\partial p(x,\sigma)}{\partial\sigma}\right|d\sigma<\epsilon k,\quad 0<x<\infty.$$

Property 9 of $p(x,t)$ in Sec. 3 has been used here.

Since the function $a(\sigma)$ given in (4.2) is continuous on $0\le\sigma\le t$ there exists a constant $k_a>0$ such that $|a(\sigma)|<k_a$, $0\le\sigma\le t$. Therefore, the second integral in (4.4) can be estimated by

$$\left|\int_{\delta}^{t}\frac{\partial p(x,\sigma)}{\partial\sigma}a(\sigma)\,d\sigma\right|<k_a\int_{\delta}^{t}\left|\frac{\partial p(x,\sigma)}{\partial\sigma}\right|d\sigma.$$

Now, $p(x,\sigma)$ takes a maximum at $\sigma=t_m>0$ and $t_m\downarrow 0$ as $x\downarrow 0$. Therefore, if $x>0$ is sufficiently small, we have $0<t_m<\delta$ and, hence, $p(x,\sigma)\downarrow$, $\delta\le\sigma$ (4, Sec. 3). Consequently,

$$\int_{\delta}^{t}\left|\frac{\partial p(x,\sigma)}{\partial\sigma}\right|d\sigma=-\int_{\delta}^{t}\frac{\partial p(x,\sigma)}{\partial\sigma}\,d\sigma<-\int_{\delta}^{\infty}\frac{\partial p(x,\sigma)}{\partial\sigma}\,d\sigma$$
$$=p(x,\delta),\quad x>0 \text{ sufficiently small},$$

where 3 of Sec. 3 has been used. Since $p(x,\delta)\downarrow 0$ as $x\downarrow 0$ (2, Sec. 3),

$$\int_{\delta}^{t}\left|\frac{\partial p(x,\sigma)}{\partial\sigma}\right|d\sigma<\epsilon \quad \text{if } x>0 \text{ is sufficiently small},$$

i.e.,

$$\left|\int_0^{t}\frac{\partial p(x,\sigma)}{\partial\sigma}a(\sigma)\,d\sigma\right|<\epsilon k_a,\quad x>0 \text{ sufficiently small}.$$

We now obtain from (4.3)

$$|z_p(x,t)-g(t)|<\epsilon(a(t)+k+k_a),\quad x>0 \text{ sufficiently small},$$

and the proof is complete.

Theorem 4.3: $x>0$, $t>0$,

$z_p(x,t)=p(x,t)*g(t)$, $g(t)\in L$ $(t_0\le t\le t_1)$,

$0\le t_0<t_1<\infty$,

$\Rightarrow z_p(x,t)$ is a boundary condition solution of $l(z)=0$ with $z_p(0+,t)=g(t)$ a.e. and $z_p(x,0+)=0$.

Proof: Let

$$a(s)=\int_0^s g(y)\,dy.$$

Then $a(s)\in C[0,\sigma]$ for every $\sigma>0$, and $a(0)=0$. We write

$$z_p(x,t)=\int_0^t p(x,t-s)\,da(s)$$

and integrate by parts. Then

$$z_p(x,t)=\int_0^t\frac{\partial p(x,t-s)}{\partial t}a(s)\,ds. \tag{4.5}$$

Let $0<\tau_1\le t\le\tau_2$. Then $\partial p(x,t-s)/\partial t\in C^{\infty}$ for $(t,s)\in[\tau_1,\tau_2]\times[0,t]$ for every $x>0$ (Theorem 3.1), and it follows that

$$\frac{\partial z_p(x,t)}{\partial t}=\int_0^t\frac{\partial^2 p(x,t-s)}{\partial t^2}a(s)\,ds, \tag{4.6}$$

[since $\partial p(x,t)/\partial t|_{t=0+}=0$].

Next, let $0<x_1\le x\le x_2$. Then $\partial p(x,t-s)/\partial t\in C^{\infty}$ for $(x,s)\in[x_1,x_2]\times[0,t]$, and we obtain from (4.5)

$$\frac{\partial z_p(x,t)}{\partial x}=\int_0^t\frac{\partial^2 p(x,t-s)}{\partial t\partial x}a(s)\,ds, \tag{4.7}$$

$$\frac{\partial^2 z_p(x,t)}{\partial x^2}=\int_0^t\frac{\partial^3 p(x,t-s)}{\partial t\partial x^2}a(s)\,ds. \tag{4.8}$$

Since $p(x,t)$ is a solution of $l(z)=0$ for $x>0$, $-\infty<t<\infty$ (Theorem 3.2), and since $l(z)$ is autonomous, it follows that $\partial p(x,t-s)/\partial t$ is also a solution of $l(z)=0$. Therefore, multiplying (4.5)–(4.8) by the appropriate coefficients $A(x)$, $B(x)$, and $C(x)$ of $l(z)$ and adding the resulting expressions, we see that $l(z_p(x,t))\equiv 0$, $x>0$, $t>0$.

The limit relations of the theorem have been established already in Theorems 4.1 and 4.2, respectively. This completes the proof.

5. THE FUNCTION $q(x,t)$

After having discussed in Sec. 3 the Green–Riemann limit function (2.3), we turn to the second limit function $\tilde{p}_2(x,t)$ given in (2.4). We define, on $x>0$, $-\infty<t<\infty$, the nonnegative function

$$q(x,t)=\begin{cases}\alpha\tilde{p}_2(x,t)=\dfrac{\alpha}{\Gamma(-\nu)}x^{-1}\varphi_0^{1-\nu}\exp(-\varphi_0), & t>0,\\ 0, & t\le 0,\end{cases} \tag{5.1}$$

where $\nu<0$ [and $\lambda=0$ in (1.1)]. The variables are again specified by (1.4) and (1.5). For $t>0$, $q(x,t)$ may be expressed in terms of the function $z_0(x,t)$ of (1.8),

$$q(x,t)=-\alpha\nu z_0(x,t),\quad t>0.$$

It should be pointed out here that for $\lambda=0$, $\nu<0$, Eq. (1.1) becomes the special Feller[6] equation. If, in addition, $\beta_2=0$, (1.1) reduces to the Kepinski[7] equation.

Theorem 5.1: $q(x,t)=\in C^{\infty}$, $x>0$, $-\infty<t<\infty$.

Proof: The same arguments used in the proof of Theorem 3.1 can be applied here. The function to be considered for $t>0$ may be expressed as

$$q(x,t)=x^{-\nu}y^{1-\nu}\exp(-xy)$$

(a constant factor has been disregarded), with

$$y=b_0^{-1}\exp\beta_2 t=b_0^{-1}+\alpha^{-1}\beta_2,\quad y\uparrow\infty \text{ as } t\downarrow 0,$$

and $dy/dt=-\alpha y^2+\beta_2 y$, so that the proof is complete.

The function $q(x,t)$ is a solution of $l(z)=0$ ($\lambda=0$, $\nu<0$) for $x>0$, $t<0$, and it has been shown that it is also a solution for $x>0$, $t>0$.

Theorem 5.2: $q(x,t)$ is a solution of $l(z)=0$ for $x>0$, $-\infty<t<\infty$.

Proof: We introduce the function

$$Q(x,t)=A(x)\frac{\partial^2 q}{\partial x^2}+B(x)\frac{\partial q}{\partial x}+C(x)q$$

with A, B, and C given by (1.1b) for $\lambda=0$. Then

$$l(q(x,t))=Q(x,t)-\frac{\partial q(x,t)}{\partial t}\equiv 0$$

for $x>0$, $t\neq 0$.

The arguments used in the proof of Theorem 3.2 are now applied to $Q(x,t)$. For $t>0$, we rewrite $Q(x,t)$ by setting $y=b_0^{-1}\exp\beta_2 t$,

$$Q(x,t)=x^{-\nu-1}y^{4-\nu}\exp(-xy)[\nu(\alpha(1+\nu)-\beta_1)y^{-2}+\beta_2(1-\nu)xy^{-2}+(2\alpha\nu-\beta_1)xy^{-1}-\beta_2x^2y^{-1}+\alpha x^2]y^{-1}.$$

(A constant factor has been disregarded.) The same technique as in the proof of Theorem 3.2 leads to the completion of the proof. The inequality we arrive at is

$$|Q(x,t)|<[(4-\nu)e^{-1}]^{4-\nu}(2x_0^{-1})^5[\nu|\alpha(1+\nu)-\beta_1|y^{-2}+\tfrac{3}{2}x_0|\beta_2|y^{-2}+\tfrac{3}{2}x_0|2\alpha\nu-\beta_1|y^{-1}+(\tfrac{3}{2}x_0)^2|\beta_2|y^{-1}+\alpha(\tfrac{3}{2}x_0)^2]y^{-1},\quad 0\le y<\infty.$$

We mention the following properties of $q(x,t)$:

(1) $q(x,t)\downarrow 0$ as $t\downarrow 0$, $x>0$.

(2) $q(x,t)\downarrow 0$ as $x\downarrow 0$, $t>0$, i.e., $q(x,t)$ is singular.

(3) $q(x,t)\to\begin{cases}[\alpha^\nu\beta_2^{1-\nu}/\Gamma(-\nu)]x^{-\nu}\exp(-\alpha^{-1}\beta_2 x)\\ \text{as } t\uparrow\infty,\ \beta_2>0,\ x>0,\\ 0 \text{ as } t\uparrow\infty,\ \beta_2\le 0,\ x>0.\end{cases}$

(4) $0<x<x_c$, $x_c=+\infty$ if $\beta_2\le 0$, $x_c=\alpha\beta_2^{-1}(1-\nu)$ if $\beta_2>0$, $\Rightarrow q(x,t)$ has exactly one maximum in $0<t<\infty$ which is located at

$$t_m=\begin{cases}\beta_2^{-1}\log\{\alpha(1-\nu)[\alpha(1-\nu)-\beta_2 x]^{-1}\}, & \beta_2\neq 0,\\ \alpha^{-1}(1-\nu)^{-1}x, & \beta_2=0,\end{cases}$$

and $t_m\downarrow 0$ as $x\downarrow 0$. For $0<x<x_c$, $q(x,t)\in\downarrow$, $0\le t\le t_m$, and $q(x,t)\in\downarrow$, $t_m\le t<\infty$. Furthermore, since $\varphi_0(t_m)\equiv 1-\nu$, $q(x,t_m)\uparrow\infty$ as $x\downarrow 0$.

(5) There exists a positive constant k_q such that

$$0<t_m q(x,t_m)\le k_q,$$

$$\begin{cases}0<x\le x_0<\alpha\beta_2^{-1}(1-\nu) & \text{if } \beta_2>0,\\ 0<x<\infty & \text{if } \beta_2\le 0.\end{cases}$$

To see this, we set $y=y(x)=\exp\beta_2 t_m$ for $\beta_2\neq 0$. Then

$$t_m q(x,t_m)=\frac{(1-\nu)^{-\nu}}{\Gamma(-\nu)}[\exp(-1+\nu)](y-1)^{-1}y\log y. \quad (5.2)$$

If $\beta_2>0$, let $0<x\le x_0<\alpha\beta_2^{-1}(1-\nu)$ so that $1<y\le y_0=\alpha(1-\nu)[\alpha(1-\nu)-\beta_2 x_0]$. Defining now the value of the function (5.2) at $y=1$ by the constant factor at the right-hand side of (5.2), we have $t_m q(x,t_m)\in C(1\le y\le y_0)$. Consequently, the function is bounded.

If $\beta_2<0$, $0<x<\infty$ implies $0<y<1$. Defining again the value of (5.2) by the same constant as before and observing that the function (5.2) is in $\downarrow$ on $0<y\le 1$, we see that it is bounded.

Finally, if $\beta_2=0$, we obtain

$$t_m q(x,t_m)\equiv\frac{(1-\nu)^{-\nu}}{\Gamma(-\nu)}\exp(-1+\nu).$$

(6) $x>0\Rightarrow$

$$\int_0^\infty q(x,t)\,dt=\begin{cases}+\infty, & \beta_2>0,\\ -\nu\Gamma(\nu;-\alpha^{-1}\beta_2 x)(-\alpha^{-1}\beta_2 x)^{-\nu}\exp(-\alpha^{-1}\beta_2 x), & \beta_2<0,\\ 1, & \beta_2=0.\end{cases} \quad (5.3)$$

To evaluate this improper integral for $\beta_2\le 0$, the substitution $\varphi_0=\sigma$ may be used.

(7) $x>0$, $\beta_2\le 0\Rightarrow$

$$Q(x)=\int_0^\infty q(x,t)\,dt\in\downarrow,\quad \lim Q(x)=1 \text{ as } x\downarrow 0. \quad (5.4)$$

The limit relation follows immediately from (5.3) (by means of de l'Hospital's rule if $\beta_2<0$) and $Q(x)$ is evidently nonincreasing for $\beta_2=0$.

Let $\beta_2<0$. We set $-\alpha^{-1}\beta_2 x=y$ and have

$$Q(x)=-\nu y^{-\nu}e^y\int_y^\infty\sigma^{\nu-1}e^{-\sigma}\,d\sigma.$$

Differentiating with respect to x and disregarding a positive constant factor, we obtain

$$\frac{dQ(x)}{dx}=y^{-\nu-1}e^y\left[(y-\nu)\int_y^\infty\sigma^{\nu-1}e^{-\sigma}\,d\sigma-y^\nu e^{-y}\right].$$

To show that the factor in brackets is negative for $0<y<\infty$, we write

$$(y-\nu)^{-1}y^\nu e^{-y}=\int_y^\infty\frac{r's-rs'}{s^2}\,d\sigma$$

with $r(\sigma)=-\sigma^\nu e^{-\sigma}$, $s(\sigma)=\sigma-\nu$. Then we compare the integrands and find

$$\sigma^{\nu-1}e^{-\sigma}<s^{-2}(r's-rs'),\quad 0<y\le\sigma,$$

since this leads to the true inequality $0<-\nu\sigma$ for $\nu<0$, $\sigma\ge y>0$.

The two properties (5.4) can be combined to give

$$0<\int_0^\infty q(x,t)\,dt\le 1,\quad 0<x<\infty,\ \beta_2\le 0.$$

(8) $x>0$, $t_0>0\Rightarrow$

$$\lim\int_0^{t_0}q(x,t)\,dt=1 \text{ as } x\downarrow 0. \quad (5.5)$$

The substitution $\varphi_0=\sigma$ leads to

$$\int_0^{t_0}q(x,t)\,dt=\begin{cases}[1/\Gamma(-\nu)]\int_{\sigma_0}^\infty\sigma^{-\nu}(\sigma-\alpha^{-1}\beta_2 x)^{-1}e^{-\sigma}\,d\sigma, & \beta_2\neq 0,\\ [1/\Gamma(-\nu)]\int_{\sigma_0}^\infty\sigma^{-\nu-1}e^{-\sigma}\,d\sigma, & \beta_2=0,\end{cases}$$

with

$$\sigma_0 = \begin{cases} \alpha^{-1}\beta_2 x(-1+\exp\beta_2 t_0)\exp\beta_2 t_0, & \beta_2 \neq 0, \\ \alpha^{-1} x t_0^{-1}, & \beta_2 = 0. \end{cases}$$

If $\beta_2 = 0$, the limit relation (5.5) is immediate.

If $\beta_2 > 0$, we have

$$0 < (\sigma - \alpha^{-1}\beta_2 x)^{-1} \leq \sigma^{-1}\exp\beta_2 t, \quad \sigma \geq \sigma_0 > \alpha^{-1}\beta_2 x.$$

We define the function

$$g(\sigma, x) = \begin{cases} \sigma^{-\nu}(\sigma - \alpha^{-1}\beta_2 x)^{-1} e^{-\sigma}, & \sigma \geq \sigma_0, \\ 0, & 0 \leq \sigma < \sigma_0, \end{cases}$$

$x > 0$, and have

$$g(\sigma, x) \leq \sigma^{-\nu-1} e^{-\sigma}\exp\beta_2 t_0, \quad 0 < \sigma$$

$$g(\sigma, 0+) = \sigma^{-\nu-1} e^{-\sigma}.$$

Lebesgue's dominated convergence theorem leads to the desired result.

Turning to $\beta_2 < 0$, we have

$$(\sigma - \alpha^{-1}\beta_2 x)^{-1} < \sigma^{-1}, \quad \sigma > 0.$$

Here, we define for $x > 0$ the function

$$g(\sigma, x) = \begin{cases} \sigma^{-\nu}(\sigma - \alpha^{-1}\beta_2 x)^{-1} e^{-\sigma} \geq \sigma_0, \\ 0, \quad 0 \leq \sigma < \sigma_0, \end{cases}$$

so that

$$g(\sigma, x) < \sigma^{-\nu-1} e^{-\sigma}, \quad 0 < \sigma,$$

$$g(\sigma, 0+) = \sigma^{-\nu-1} e^{-\sigma}.$$

Again, Lebesgue's dominated convergence theorem establishes the limit relation (5.5).

(9) There exists a constant $k > 0$ such that

$$\int_0^\tau t \left| \frac{\partial q(x,t)}{\partial t} \right| dt < k, \quad 0 < x < x_0,$$

where

$$\tau = \begin{cases} t_0 > 0, & \beta_2 > 0, \\ +\infty, & \beta_2 \leq 0, \end{cases}$$

$$x_0 = \begin{cases} (1-\nu) b_0(t_0)\exp(-\beta_2 t_0), & \beta_2 > 0, \\ +\infty, & \beta_2 \leq 0, \end{cases}$$

and $k = k(t_0)$ for $\beta_2 > 0$.

We consider the case $\beta_2 > 0$ first. The condition $0 < x < x_0$ implies $t_m(x) < t_0$ where t_m is defined under 4. Then for $0 < x < x_0$,

$$\frac{\partial q(x,t)}{\partial t} \begin{cases} \geq 0, & 0 \leq t \leq t_m(x), \\ \leq 0, & t_m(x) \leq t \leq t_0. \end{cases}$$

Therefore, splitting the integral from 0 to t_0 into two, one from 0 to $t_m(x)$, the other from $t_m(x)$ to t_0, we obtain by partial integration

$$\int_0^{t_0} t \left| \frac{\partial q(x,t)}{\partial t} \right| dt < 2t_m q(x, t_m) + t_0 q(x, t_0) + \int_0^{t_0} q(x,t)\, dt, \quad 0 < x < x_0. \tag{5.6}$$

The first term is bounded by $2k_q$ (see 5). Furthermore,

$$0 < q(x, t_0) \leq \max_{0<x<\infty} q(x, t_0)$$

$$= \frac{\alpha}{(-\nu)^\nu \Gamma(-\nu)} b_0^{-1}(t_0)\exp\beta_2 t_0 = k_1(t_0).$$

Finally,

$$Q < \int_0^{t_0} q(x,t)\, dt \leq \exp\beta_2 t_0 = k_2(t_0).$$

Consequently, the right-hand side of (5.6) is not greater than

$$k = k(t_0) = 2k_q + k_1(t_0) + k_2(t_0) \text{ for } 0 < x < x_0.$$

If $\beta_2 \leq 0$, we may follow the arguments used under 9 of Sec. 3 to see that $k = 2k_q + 1$.

6. BOUNDARY CONDITION SOLUTIONS BY THE q-CONVOLUTION

Using the function $q(x,t)$ defined in (5.1), we introduce the unilateral convolution

$$z_q(x,t) = q(x,t) * g(t) = \int_0^t q(x, t-s) g(s)\, ds, \quad x > 0, \ t > 0. \tag{6.1}$$

The discussions of this section parallel those of Sec. 4.

Theorem 6.1: $x > 0$, $t > 0$,

$z_q(x,t) = q(x,t) * g(t)$, $g(t) \in \mathcal{L}$ $(t_0 \leq t \leq t_1)$,

$0 \leq t_0 < t_1 < \infty$, $\Rightarrow z_q(x,t)$ exists and $z_q(x, 0+) = 0$.

Proof: If we replace p by q, the proof of Theorem 4.1 carries over verbatim.

Theorem 6.2: $x > 0$, $t > 0$,

$z_q(x,t) = q(x,t) * g(t)$, $g(t) \in \mathcal{L}$ $(t_0 \leq t \leq t_1)$

$0 \leq t_0 < t_1 < \infty$ $\Rightarrow z_q(0+, t) = g(t)$ a.e.

Proof: With p replaced by q, the various steps of the proof of Theorem 4.2 carry over verbatim to the present situation up to and including identity (4.4), i.e.,

$$\int_0^t \frac{\partial q(x,\sigma)}{\partial \sigma} a(\sigma)\, d\sigma = \int_0^\delta + \int_\delta^t, \tag{6.2}$$

and the first integral at the right-hand side of (6.2) can be estimated as

$$\left| \int_0^\delta \frac{\partial q(x,\sigma)}{\partial \sigma} a(\sigma)\, d\sigma \right| < \epsilon \int_0^\delta \sigma \left| \frac{\partial q(x,\sigma)}{\partial \sigma} \right| d\sigma. \tag{6.3}$$

Let $\beta_2 > 0$. We have

$$\int_0^\delta \sigma \left| \frac{\partial q(x,\sigma)}{\partial \sigma} \right| d\sigma < \int_0^t \sigma \left| \frac{\partial q(x,\sigma)}{\partial \sigma} \right| d\sigma. \tag{6.4}$$

According to 9, Sec. 5, there exists a positive constant $k = k(t)$, such that the integral at the right-hand side of (6.4) is less than k if $0 < x < x_0 = x_0(t) = (1-\nu) b_0(t) \times \exp(-\beta_2 t)$. Thus, it follows from (6.3) that

$$\left| \int_0^\delta \frac{\partial q(x,\sigma)}{\partial \sigma} a(\sigma)\, d\sigma \right| < \epsilon k, \quad 0 < x < x_0. \tag{6.5}$$

Using again 9 of Sec. 5 in the case $\beta_2 \leq 0$, we obtain at once (6.5), now for arbitrary positive x. Consequently, inequality (6.5) holds for arbitrary β_2 provided $x > 0$ is sufficiently small.

We look now at the second integral at the right-hand side of (6.2). Using the same notation as in the corresponding part of the proof of Theorem 4.2, we arrive at

$$\left| \int_0^t \frac{\partial q(x,\sigma)}{\partial \sigma} a(\sigma)\, d\sigma \right| < k_a \int_\delta^t \left| \frac{\partial q(x,\sigma)}{\partial \sigma} \right| d\sigma.$$

Now, $q(x,\sigma)$ takes a maximum at $\sigma = t_m$ [with $0 < x < (1-\nu)\alpha\beta_2^{-1}$ if $\beta_2 > 0$, see 4, Sec. 5] and $t_m \downarrow 0$ as $x \downarrow 0$. Consequently, if $x > 0$ is sufficiently small, we have $0 < t_m < \delta$ and, hence, $q(x,\sigma) \in \downarrow$ for $\delta \le \sigma \le t$. Therefore,

$$\int_0^t \left| \frac{\partial q(x,\sigma)}{\partial \sigma} \right| d\sigma = -\int_0^t \frac{\partial q(x,\sigma)}{\partial \sigma} d\sigma = q(x,\delta) - q(x,t). \tag{6.6}$$

Since $q(x,t_0) \downarrow 0$ as $x \downarrow 0$ for every $t_0 > 0$, it follows from (6.6) that

$$\int_\delta^t \left| \frac{\partial q(x,\sigma)}{\partial \sigma} \right| d\sigma < \epsilon, \quad x > 0 \text{ sufficiently small,}$$

and, consequently,

$$\left| \int_\delta^t \frac{\partial q(x,\sigma)}{\partial \sigma} a(\sigma)\, d\sigma \right| < \epsilon k_a, \quad x > 0 \text{ sufficiently small.}$$

Therefore, we finally arrive at the inequality

$$|z_q(x,t) - g(t)| < \epsilon(a(t) + k + k_a),$$

$x > 0$ sufficiently small,

which completes the proof.

Furthermore, the following theorem holds for the convolution (6.1).

Theorem 6.3: $x > 0$, $t > 0$,

$z_q(x,t) = q(x,t) * g(t)$, $g(t) \in \mathcal{L}$ $(t_0 \le t \le t_1)$,

$0 \le t_0 < t_1 < \infty$,

$\Rightarrow z_q(x,t)$ is a boundary condition solution of $l(z) = 0$ with $z_q(0+,t) = g(t)$ a.e. and $z_q(x,0+) = 0$.

Proof: With p replaced by q, the proof of Theorem 4.3 carries over verbatim to the present case.

7. INITIAL AND BOUNDARY CONDITION SOLUTIONS

We consider now the function

$$z_*(x,t) = \int_0^\infty v^*(x,t;y) f(y)\, dy, \quad x > 0, \quad 0 < t < t_0,$$

given by (1.2) under the conditions on $f(y)$ mentioned there and with v^* specified in (1.3)–(1.6). We repeat the main result of Ref. 1.

Theorem 7.1: With the kernel $v^*(x,t;y)$ given by (1.3)–(1.6) and under the hypotheses on $f(y)$ listed in connection with (1.2), the singular integral

$$z_*(x,t) = \int_0^\infty v^*(x,t;y)\, f(y)\, dy$$

is an initial condition solution of $l(z)$ on the half-strip $x > 0$, $0 < t < t_0$, and

$z_*(x,0+) = f(x)$ a.e.

Assume now that $f: (0,\infty) \to \mathbb{R}$ is measurable and bounded. Theorem 7.1 remains true in this special case for $x > 0$, $t > 0$. In addition, we prove

Theorem 7.2: $x > 0$, $t > 0$,

(a) $\lambda < 0$, $\nu > 0$, $\bar{\mathcal{L}}_\nu(r_0) = I_\nu(r_0)$ in (1.3),

(b) $\lambda + \nu(1-\lambda) < 0$, $\nu < 1$, $\bar{\mathcal{L}}_\nu(r_0) = I_{-\nu}(r_0)$ in (1.3),

$f(y)$ measurable and bounded $\Rightarrow$

$$z_*(0+,t) = \lim_{x \downarrow 0} \int_0^\infty v^*(x,t;y) f(y)\, dy = 0.$$

Proof: If $|f(y)| < f_0$, $0 < x < \infty$, then

$$|z_*(x,t)| < f_0 \int_0^\infty v^*(x,t;y)\, dy$$

$$= c_1 \varphi_0^\lambda \Phi(\nu + (1-\lambda)^{-1}, 1+\nu; \varphi_0^{1-\lambda}) \exp(-\varphi_0^{1-\lambda} + \beta_2 t), \tag{7.1}$$

$$= c_2 \varphi_0^{-\lambda-\nu(1-\lambda)} \Phi((1-\lambda)^{-1}, 1-\nu; \varphi_0^{1-\lambda}) \exp(-\varphi_0^{1-\lambda} + \beta_2 t). \tag{7.2}$$

Estimate (7.1) holds for case (a), (7.2) for (b). Furthermore, c_1 and c_2 are certain positive constants and Φ is the confluent hypergeometric function. The expressions for the integral of v^* can be found in Ref. 8, 6.643.2 (see also Ref. 1, Proof of Theorem 5.1). The limit relation of the theorem is now obvious.

From Theorems 4.3, 7.1, and 7.2(a), we now obtain

Theorem 7.3: $x > 0$, $t > 0$, $\lambda < 0$,

$\nu = -\lambda(1-\lambda)^{-1}$, $\bar{\mathcal{L}}_\nu(r_0) = I_\nu(r_0)$ in (1.3),

$f(x)$ measurable and bounded on $(0,\infty)$,

$g(t) \in \mathcal{L}$ $(t_0 \le t \le t_1)$, $0 \le t_0 < t_1 < \infty$ $\Rightarrow$

$$z(x,t) = \int_0^\infty v^*(x,t;y) f(y)\, dy + \int_0^t p(x,t-s) g(s)\, ds$$

is an initial and boundary condition solution of $l(z) = 0$ with

$z(x,0+) = f(x)$ a.e., $z(0+,t) = g(t)$ a.e.

We also combine Theorems 6.3, 7.1, and 7.2(b).

Theorem 7.4: $x > 0$, $t > 0$, $\lambda = 0$, $\nu < 0$,

$\bar{\mathcal{L}}_\nu(r_0) = I_{-\nu}(r_0)$ in (1.3),

$f(x)$ measurable and bounded on $(0,\infty)$,

$g(t) \in \mathcal{L}$ $(t_0 \le t \le t_1)$, $0 \le t_0 < t_1 < \infty$ $\Rightarrow$

$$z(x,t) = \int_0^\infty v^*(x,t;y) f(y)\, dy + \int_0^t q(x,t-s) g(s)\, ds$$

is an initial and boundary condition solution of $l(z) = 0$ with

$z(x,0+) = f(x)$ a.e., $z(0+,t) = g(t)$ a.e.

[1]S.H. Lehnigk, "Initial condition solutions of the generalized Feller equation," accepted for publication, J. Appl. Math. Phys. (ZAMP).
[2]S.H. Lehnigk, "A class of conservative diffusion processes with delta function initial conditions," J. Math. Phys. **17**, 973—76 (1976).
[3]S.H. Legnigk, "Maxwell and Wien processes as special cases of the generalized Feller diffusion process," J. Math. Phys. **18**, 104—05 (1977).
[4]G. Doetsch, 'Der lineare Wärmeleiter mit verschwindender Anfangstemperatur. Die allgemeinste Lösung und die Frage der Eindeutigkeit," Math. Z. **22**, 293—306 (1925).
[5]D.V. Widder, *The Heat Equation* (Academic, New York, 1975).
[6]W. Feller, "Two singular diffusion problems," Ann. Math. **54**, 173—82 (1951).
[7]S. Kepinski, "Uber die Differentialgleichung $\partial^2 z/\partial x^2 + [(m+1)/x]\partial z/\partial x - [n/x]\partial z/\partial t = 0$," Math. Ann. **61**, 397—405 (1905).
[8]I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), 4th ed.

The structure of the many-body wavefunction for scattering[a)]

M. L'Huillier

Institut de Physique Nucléaire, Division de Physique Théorique, 91406 Orsay, France

Edward F. Redish

Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

P. C. Tandy

Research School of Physical Sciences, The Australian National University, Canberra, A.C.T., 2600 Australia

(Received 2 December 1977)

We show that the scattered part of the many-body wavefunction initiated by two incoming clusters is given by a fully connected operator acting on the initial channel state. The structure of this operator suggests a division of the full wavefunction into two-cluster components. A set of coupled equations in both the differential and integral form is then derived for these components. These equations have structure and properties similar to the three-body equations of Faddeev. We demonstrate that each component has outgoing waves in a unique two-cluster partition. The transition amplitude for any final arrangement can therefore be extracted directly from the outgoing waves in the relevant components.

I. INTRODUCTION

The essential difficulty in the N-body scattering problem is that there can be different arrangements of the particles. In the asymptotic region of the many-particle coordinate space, the "free" propagation is therefore not governed by a single channel Hamiltonian. As a result, it is hard to specify boundary conditions. If one uses a dynamical equation which describes the asymptotic propagation in terms of one of the channel Hamiltonians, real calculations will have trouble producing outgoing waves in the other arrangements.

These difficulties are overcome in the case of three particles by the method of Faddeev.[1-3] The method can be realized in equations either for wavefunctions or transition operators, in momentum or coordinate representations, and via differential or integral equations.

In the nonrelativistic many-body scattering problem most recent work has emphasized the connectivity structure of the kernel of an integral equation for transition operators.[4-15] This eliminates the troublesome disconnected diagrams,[4] permits some iterate of the kernel to be compact, and thereby offers the possibility of solutions by standard numerical techniques. An alternative view focuses on the structure of the outgoing waves in coordinate space. This provides a different perspective on the problem and may yield new insights for developing approximate methods.

In this paper we divide the N-body scattering wavefunction into components and find equations for them. These equations reduce to Faddeev's for the case $N=3$, have the Faddeev structure for all N, and solve the difficulties of the asymptotic behavior of the N-body wavefunction. In Sec. II we briefly review the Faddeev treatment of the three-body wavefunction and pick out the essential features. In Sec. III we define our notation for the N-body problem and recall a number of basic results. In Sec. IV we demonstrate that the scattered part of the N-body wavefunction for a two-cluster initial state is equal to a completely connected operator acting on the initial channel state. The structure of this operator suggests a decomposition of the wavefunction into two-cluster components. This is done in Sec. V where equations for these components are derived and their structure is discussed. A summary and conclusions are presented in Sec. VI.

II. THE THREE-BODY CASE

We begin by reviewing briefly Faddeev's method to stress its essential characteristics. Let ψ_i be the exact wavefunction for particle i incident on the other pair of particles in their bound state. (For simplicity of presentation we assume each pair of particles has a single bound state.) We label the pair of particles *not* including particle i as the "i-pair." The wavefunction ψ_i satisfies

$$(E-H)\psi_i = 0. \tag{1}$$

We introduce the channel Hamiltonians

$$H_i = H_0 + V_i, \tag{2}$$

where V_i is the interaction of the i-pair and H_0 is the kinetic energy of all the particles. The initial channel state ϕ_i satisfies

$$H_i\phi_i = E\phi_i. \tag{3}$$

The wavefunction ψ_i then satisfies the set of Lippmann–Schwinger (LS) integral equations

$$\psi_i = \phi_i\delta_{ij} + G_jV^j\psi_i, \quad j=1,2,3, \tag{4}$$

where

$$V^j = H - H_j \tag{5}$$

is the residual interaction and

$$G_j = (E - H_j + i\epsilon)^{-1} \tag{6}$$

[a)]Research supported in part by the U.S. ERDA.

 0022-2488/78/1906-1276$1.00

is the channel Green function. It is understood that the limit $\epsilon \to 0+$ is to be taken at the end of the calculation. All three of the equations (4) are necessary to specify that there are no incoming waves in rearrangement channels.[16] No single one suffices to specify ψ_i uniquely since outgoing waves in one channel may be superposed to give an incoming wave in a rearrangement channel. Furthermore, it would be difficult to describe outgoing waves of all three types with a single Green function.

Some insight into these problems can be gained by considering the spatial extent of the source term for Eq. (4). Let $S_j^{\mathrm{LS}} = V^j\psi_i = (V_i + V_k)\psi_i$ where (i, j, k) is some permutation of (1, 2, 3). We denote the relative separation of the j-pair by $\mathbf{r}_j$, and the separation of particle j from the center of mass of the j-pair by $\mathbf{R}_j$. The positions of the three particles in their center of mass system can be described by any one of the three sets $(\mathbf{r}_i, \mathbf{R}_i)$, $i = 1, 2, 3$. Take a fixed j and consider the asymptotic behavior of $S_j^{\mathrm{LS}}(\mathbf{r}, \mathbf{R})$ in each of the three directions $r_l < c$, $R_l \to \infty$ where $l = 1, 2, 3$, and c is a constant specifying the range of a pair interaction. When $l = j$, S_j^{LS} decays rapidly due to the potentials. However, when $l \neq j$, one of the potentials in V^j remains constant and S_j^{LS} falls off like the wavefunction, i.e. $S_j^{\mathrm{LS}} \sim \exp(ip_lR_l)/R_l$. Thus the source term is not confined to a finite volume. Since the angular momentum barrier term of H_0 depends on the inverse square of distance, the source term should fall off faster than an inverse square law before the boundary condition on the wavefunction can be determined from that of the Green function G_j.

These problems are overcome by using all three of the equations (4). These three separate equations, which must be satisfied simultaneously, can be combined into a set of coupled equations by dividing ψ_i into three components. This can be done by multiplying Eq. (4) by $(1 - G_0V_j)$ to get

$$\psi_i = G_0V\psi_i. \tag{7}$$

Since we want to divide ψ_i into parts propagated asymptotically by their own channel Hamiltonian, we define

$$\psi_i^{(j)} = G_0V_j\psi_i, \tag{8}$$

with

$$\psi_i = \sum_j \psi_i^{(j)}. \tag{9}$$

Since V_j is localized to the region where the particles of the j-pair are close together, we expect that $\psi_i^{(j)}$ propagates asymptotically only by G_j. To see this we substitute Eq. (9) into Eq. (8) and multiply by G_0^{-1} to get the set of coupled equations

$$(E - H_0 - V_j)\psi_i^{(j)} = V_j\sum_k \bar{\delta}_{jk}\psi_i^{(k)}, \tag{10}$$

where $\bar{\delta}_{jk}$ is one if $j \neq k$ and 0 if $j = k$.

These are the Faddeev equations for the wavefunction components $\psi_i^{(j)}$. The advantage in this approach is that the source terms are confined. Consider the source $S_j^F = V_j\sum_k \bar{\delta}_{jk}\psi_i^{(k)}$ of Eq. (10). From Eq. (8), $\psi_i^{(k)}$ is that part of ψ_i in which the k-pair interacts last. The potential V_j requires both particles of the j-pair to be within range of their interaction. Since the $\bar{\delta}_{jk}$ requires the j- and k-pairs to be different, S_j^F is made up from terms in which each of the three particles is experiencing an interaction. This contrasts with S_j^{LS} which depends on the full wavefunction and so contains terms in which a given pair experiences repeated interactions.

As a result, S_j^F is confined to a finite volume, and in the asymptotic region the component $\psi_i^{(j)}$ satisfies

$$(E - H_j)\psi_i^{(j)} = 0. \tag{11}$$

The full wavefunction ψ_i has therefore been divided into components $\psi_i^{(j)}$ each of which is propagated everywhere in the asymptotic region by its own channel Hamiltonian. Writing the set (10) in integral form yields

$$\psi_i^{(j)} = \delta_{ij}\phi_i + G_jV_j\sum_k \bar{\delta}_{jk}\psi_i^{(k)}. \tag{12}$$

It can be shown[2] that these equations suffice to specify a unique solution. The details of the asymptotic behavior of the components $\psi_i^{(j)}$ is readily deduced from the singularity structure in momentum space.[17] We use these results to illustrate the confinement of the sources. In the directions $r_l < c$, $R_l \to \infty$, described above, S_j^F is well confined for $l \neq j$ due to the potential V_j. When $l = j$, this potential is constant and S_j^F behaves in this direction like the wavefunction components $\sum_k \bar{\delta}_{jk}\psi_i^{(k)}$.

Consider firstly the asymptotic behavior when a pair remains bound. Only the k-pair can remain bound in the component $\psi_i^{(k)}$. However in the direction $R_j \to \infty$, the bound state wavefunctions of the pairs other than j decay exponentially. The leading asymptotic behavior in this direction is provided by the breakup contribution to $\psi_i^{(k)}(\mathbf{r}_k, \mathbf{R}_k)$. Both of the coordinates $\mathbf{r}_k$, $\mathbf{R}_k$ for $k \neq j$, are linear combinations of $\mathbf{r}_j$ and $\mathbf{R}_j$. Hence in the direction $r_j < c$, $R_j \to \infty$, we have $r_k \to \infty$ and $R_k \to \infty$ while $r_k/R_k \to$ constant. The components $\sum_k \bar{\delta}_{jk}\psi_i^{(k)}$ and therefore the source S_j^F, fall off according to the power law $S_j^F \sim (R_j)^{-5/2}$ which is faster than the angular momentum barrier. Thus in the integral equation (12) the boundary conditions on the wavefunction are well determined by those of the Green functions. Because of the power law behavior the volume of confinement of the sources can be quite large. Such long range effects are necessarily present in the three-body problem when the breakup channel is open.

In summary, the crucial properties that make the Eqs. (10) and (12) a solution of the difficulties are:

(i) The wavefunction is divided into parts $\psi_i^{(j)}$ which satisfy Eq. (11) anywhere in the asymptotic region. Therefore, $\psi_i^{(j)}$ may only have outgoing waves having a bound j-pair or three-body breakup. There is no mixing in of bound states in other arrangements. Only the diagonal term $\psi_i^{(i)}$ has the incoming wave.

This property follows from:

(ii) The source (coupling) term in Eq. (10) is confined in the sense that it falls off with distance faster than the angular momentum barrier. This confinement is an expression of the fact that the iterated kernel in Eq. (12) is completely connected.

III. NOTATION AND BASIC RELATIONS

We consider a system of N distinguishable particles

interacting by two-body potentials. A division of the particles into groups is called a partition and is labeled by Latin letters $a, b, c, \cdots$. The number of clusters in partition a is n_a. The two-cluster partitions play a special role in our formulation so we label them distinctively by Greek letters $\alpha, \beta, \gamma, \cdots$. We restrict our analysis to states initiated by an incoming wave of two bound clusters of some partition β. We write this state as ψ_β suppressing the quantum numbers associated with the particular bound states of the clusters. This state satisfies the Schrödinger equation

$$(E - H)\psi_\beta = 0 \tag{13}$$

and the set of LS equations

$$\psi_\beta = \delta_{a\beta}\phi_\beta + G_a V^a \psi_\beta, \tag{14}$$

where ϕ_β is the incoming state to be defined below.

We use the definitions that H_0 is the total kinetic energy operator for the N particles, V_b is the sum of the two-body interactions internal to the clusters of b, and V^b is the sum of the interactions between the particles in different clusters of b. The partition Hamiltonians are

$$H_b = H_0 + V_b, \tag{15}$$

and the full Hamiltonian satisfies

$$H = H_b + V^b, \tag{16}$$

for any partition b. The full Green function is

$$G(E + i\epsilon) = (E - H + i\epsilon)^{-1}, \tag{17}$$

and the partition Green functions are

$$G_b(E + i\epsilon) = (E - H_b + i\epsilon)^{-1}. \tag{18}$$

The incoming state ϕ_β is the product of the internal wavefunctions of the clusters with a plane wave for their relative motion. It satisfies

$$(E - H_\beta)\phi_\beta = 0. \tag{19}$$

We observe that this implies

$$G_0 V_\beta \phi_\beta = \phi_\beta \tag{20}$$

when the energy of the operator matches the energy of the state, i.e., on shell.

For an arbitrary partition a the Green function satisfies a resolvent equation

$$G = G_a + G_a V^a G = G_a + G V^a G_a. \tag{21}$$

The exact wavefunction is defined by[18]

$$\psi_\beta = \lim_{\epsilon \to 0} i\epsilon G(E + i\epsilon)\phi_\beta. \tag{22}$$

We use Eqs. (18) and (19) to express the wavefunction as

$$\psi_\beta = G G_\beta^{-1}\phi_\beta. \tag{23}$$

This wavefunction is actually a function of the parameter ϵ, but this dependence is not made explicit. In this equation and in other equations involving Green functions considerable care must be taken to maintain the $i\epsilon$ not equal to zero until all formal manipulations have been carried out. (See for example the discussion in Ref. 19 and in Appendix A below).

We now introduce the transition operators

$$T^{ab} = V^a + V^a G V^b. \tag{24}$$

The matrix elements of these operators between on-shell channel states give the scattering probability amplitudes.[18] Using Eq. (21) these operators can be expressed in the convenient form

$$T^{ab} = V^a G G_b^{-1}. \tag{25}$$

This implies that

$$V^a \psi_\beta = T^{a\beta}\phi_\beta. \tag{26}$$

In the limit $\epsilon \to 0$, we may write

$$\psi_\beta = G_0 T^{(\beta}\phi_\beta, \tag{27}$$

where the relation

$$G_0 V \psi_\beta = \psi_\beta, \tag{28}$$

which holds in the same limit, has been used.

IV. THE LINKED CLUSTER THEOREM

We can now transfer the information which has been learned in the past few years about the structure of the transition operators to the wavefunction via Eq. (27). In particular, we eventually have applications to nuclear, atomic, and molecular problems in mind. Since highly clustered systems dominate most of reaction theory, we apply the Bencze–Redish–Sloan (BRS) equations.[9–11] These equations relate the two-cluster to two-cluster transition operators, so highly clustered final and intermediate states are easily accessible.[20]

The BRS equations state that the transition operators defined by Eq. (24) satisfy[11]

$$T^{\alpha\beta}\phi_\beta = V_\beta^\alpha \phi_\beta + \sum_\gamma K_\gamma^\alpha G_0 T^{\gamma\beta}\phi_\beta. \tag{29}$$

The two-cluster to two-cluster operators are coupled in a set of $2^{N-1} - 1$ equations, while the breakup amplitudes are given by quadratures. The potential V_β^α is the sum of all two-body potentials which are both internal to β and external to α. The kernel operator K_γ^α can be described[20] as the sum of all Weinberg graphs[4] having connectivity γ and ending (on the left) with an interaction external to α.

In the BRS Eqs. (29), the inhomogeneous term has been simplified by omitting contributions which vanish in the limit $\epsilon \to 0$. In Ref. 19 it is demonstrated that the omitted parts of the inhomogeneous term do not contribute to the physical solutions for the transition amplitudes. The omitted parts also do not contribute to the equations for the physical wavefunction that are derived in the following section. For simplicity of presentation, we work with Eq. (29) carrying $\epsilon \neq 0$ and justify this procedure in Appendix A.

Inserting the BRS equations (29) into the expression (27) gives

$$\psi_\beta = \phi_\beta + \sum_\gamma G_0 K_\gamma G_0 T^{\gamma\beta}\phi_\beta, \tag{30}$$

where we have written K_γ for K_γ^0 and Eq. (20) has been used. We refer to this result as the *linked cluster theorem* dues to its characterization in terms of graphs.

Recalling Eq. (24), we see that, apart from the first term, $T^{\gamma\beta}$ is the sum of all graphs beginning with an interaction external to β and ending with an interaction external to γ.

This means that a major part of the scattered wavefunction, $\psi_\beta^s = \psi_\beta - \phi_\beta$, may be characterized as the sum of all completely connected graphs beginning with an interaction external to β acting on the initial state ϕ_β. To be more precise, let us introduce

$$V^{\gamma\beta} = V^\gamma - V_\beta^\gamma \tag{31}$$

which is the sum of all two-body interactions external to both of the partitions γ and β.

Then the operator $T^{\gamma\beta}$ can be written

$$T^{\gamma\beta} = V_\beta^\gamma + \tau^{\gamma\beta}, \tag{32}$$

where

$$\tau^{\gamma\beta} = V^{\gamma\beta} + V^\gamma G V^\beta \tag{33}$$

is the sum of all Weinberg graphs beginning with an interaction external to β and ending with one external to γ. Inserting Eq. (32) into Eq. (30) we get the following representation of the scattered wave

$$\psi_\beta^s = \psi_\beta - \phi_\beta = \sum_\gamma [G_0 K_\gamma G_0 \tau^{\gamma\beta}]\phi_\beta + \sum_\gamma [G_0 K_\gamma G_0 V_\beta^\gamma]\phi_\beta. \tag{34}$$

We define the operators

$$K^{0\beta} = \sum_\gamma K_\gamma G_0 \tau^{\gamma\beta}, \tag{35}$$

$$W_\beta = \sum_\gamma K_\gamma G_0 V_\beta^\gamma. \tag{36}$$

From the graphical characterizations of K_γ and $\tau^{\gamma\beta}$ we see that $K^{0\beta}$ is the sum of all completely connected graphs starting with an interaction external to β. The operator W_β is the sum of all completely connected graphs which begin with an interaction internal to β but which are irreducible in the sense of Weinberg,[4] that is, they become disconnected when the first interaction is removed. With these definitions the linked cluster theorem, Eq. (30), can be expressed as

$$\psi_\beta^s = G_0 K^{0\beta}\phi_\beta + G_0 W_\beta \phi_\beta. \tag{37}$$

We note that a somewhat similar structure is demonstrated in the work of Rosenberg.[5]

V. DIVISION OF THE WAVEFUNCTION INTO TWO-CLUSTER COMPONENTS

The structure of the linked-cluster theorem, Eq. (30), suggests a division of the wavefunction into components labeled by the two-cluster partitions. We divide ψ_β as follows,

$$\psi_\beta = \sum_\gamma \psi_\beta^{(\gamma)}, \tag{38}$$

where

$$\psi_\beta^{(\gamma)} = \delta_{\gamma\beta}\phi_\beta + G_0 K_\gamma G_0 T^{\gamma\beta}\phi_\beta. \tag{39}$$

Since K_γ is the sum of all possible Weinberg graphs of connectivity γ, it contains all those interactions necessary to combine the particles of the system into the two bound clusters of the partition γ. This is done by an infinite number of terms adding to contribute through K_γ a primary singularity at the end of the graph.[21] This generates an outgoing wave in the coordinate space representation.[17] Any infinite sequence appearing in the middle of the graph which is associated with a two-cluster partition $\alpha \neq \gamma$ does not yield a primary singularity. The final part of the graph having connectivity γ will have at least one interaction external to α occuring after the singular part. The intermediate integration in the relative coordinate associated with this potential weakens the singularity enough to prevent it from producing primary singularities in the final state.[22]

As a result, the only outgoing waves present in a component $\psi_\beta^{(\gamma)}$ are of the γ type. This means that it may have outgoing waves only in the two-cluster channels of the partition γ, and only those breakup waves which can be obtained by breaking the clusters of the partition γ. The outgoing waves in a channel of partition a having more than two clusters are in all those components $\psi_\beta^{(\gamma)}$ for which the partition γ can be obtained by combining two or more clusters of a. (To denote this relation we write $\gamma \supset a$.)

We now derive the equations satisfied by the components. We do this by expressing the $T^{\gamma\beta}\phi_\beta$ appearing in Eq. (39) in terms of the full wavefunction ψ_β. The division (38) is then used to produce coupled equations of the Faddeev type for the components.

To express $T^{\gamma\beta}\phi_\beta$ in terms of ψ_β we use Eq. (23) and the resolvent equation (21) for the partition γ. This gives

$$\psi_\beta = G_\gamma G_\beta^{-1}\phi_\beta + G_\gamma T^{\gamma\beta}\phi_\beta. \tag{40}$$

Equation (25) has also been used. Solving for $T^{\gamma\beta}\phi_\beta$ gives

$$T^{\gamma\beta}\phi_\beta = G_\gamma^{-1}\psi_\beta - G_\beta^{-1}\phi_\beta. \tag{41}$$

Inserting this in Eq. (39) gives

$$\psi_\beta^{(\gamma)} = (\delta_{\gamma\beta} - G_0 K_\gamma G_0 G_\beta^{-1})\phi_\beta + G_0 K_\gamma G_0 G_\gamma^{-1}\psi_\beta. \tag{42}$$

[The structure of the inhomogeneous term appears as it does here because contributions which vanish in the limit $\epsilon \to 0$ have been omitted in obtaining the simplified structure of the first term of Eq. (39). The validity of these subtle limiting procedures is shown in Appendix A.]

The inhomogeneous term of Eq. (42) simplifies on-shell. We apply the anticluster expansion[20] for the operator K_γ. This is the inversion of the Yakubovskii cluster expansion[6] and takes the form

$$K_\gamma G_0 = \sum_{(\gamma \supseteq)a} N(\gamma, a) V_a G_a. \tag{43}$$

The numerical coefficients $N(\gamma, a)$ guarantee cancellation of the parts having connectivities other than γ. Labels bracketed under the summation sign are not summed over. This yields

$$G_0 K_\gamma G_0 G_\beta^{-1}\phi_\beta = \sum_{(\gamma \supseteq)a} G_0 N(\gamma, a) V_a G_a G_\beta^{-1}\phi_\beta. \tag{44}$$

Applying the Lippmann identity[23]

$$G_a G_\beta^{-1}\phi_\beta = \delta_{a\beta}\phi_\beta, \quad a \not\supset \beta, \tag{45}$$

which holds in the limit $\epsilon \to 0$, Eq. (44) becomes

$$G_0 K_\gamma G_0 G_\beta^{-1}\phi_\beta = G_0 V_\beta \phi_\beta \delta_{\gamma\beta}, \tag{46}$$

where we have used the fact that $N(\gamma,\gamma)=1$.

By Eq. (20) this term cancels the rest of the inhomogeneity in Eq. (42) giving

$$\psi_\beta^{(\gamma)} = G_0 K_\gamma G_0 G_\gamma^{-1}\psi_\beta. \tag{47}$$

It is convenient to introduce the operator V_γ by the relation

$$V_\gamma G_\gamma = K_\gamma G_0. \tag{48}$$

We refer to V_γ as the *irreducible γ-connected potential*. Since G_γ contains the term G_0, V_γ has the same connectivity as K_γ. It is the sum of all γ-connected graphs which become more disconnected when their rightmost interaction is removed. In the case $N=3$, partition γ specifies a pair, K_γ is the two-body T matrix t_γ, and V_γ is simply the pair interaction.

For $N>3$, an explicit expression for V_γ can be obtained in terms of strings of the fundamental potentials and Green functions. Working from the corresponding relation[10] for K_γ and using Eq. (48) gives

$$V_\gamma = \sum_{a_{N-1}\subset a_{N-2}\subset\cdots\subset a_3 \subset (\gamma)} V_{a_{N-1}} G_{a_{N-1}} V_{a_{N-2}}^{a_{N-1}} G_{a_{N-2}} \cdots G_{a_3} V_\gamma^{a_3}, \tag{49}$$

where we have written a_m to indicate an m-cluster partition.

To illustrate the structure of this operator, we display it for the particular example of $N=4$, $\gamma=(12)(34)$. In this case the sum reduces to

$$V_\gamma = \sum_{a_3 \subset (\gamma)} V_{a_3} G_{a_3} V_\gamma^{a_3}. \tag{50}$$

This sum contains the two terms for $a_3=(1)(2)(34)$ and $a_3=(12)(3)(4)$ giving

$$V_\gamma = t_{12} G_0 v_{34} + t_{34} G_0 v_{12}, \tag{51}$$

where v_{ij} is the interaction of the pair ij and the relation $v_{ij}G_{ij} = t_{ij}G_0$ has been used. The connectivity and irreducibility structure can easily be seen from the graphical representation of Eq. (51) as shown in Fig. 1.

Returning to the general case, using Eq. (48), Eq. (47) becomes

$$\psi_\beta^{(\gamma)} = G_0 V_\gamma \psi_\beta. \tag{52}$$

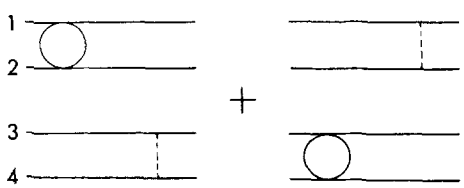


FIG. 1. Graphical representation of the irreducible two-cluster potential operator V_γ for the case $N=4$ and $\gamma=(12)(34)$. Circles indicate two-body T matrices and dotted lines indicate two-body potentials.

This is completely parallel to the three-body result, Eq. (8). Multiplying by G_0^{-1}, dividing ψ_β into parts by Eq. (38), and bringing the diagonal term to the left gives the equation

$$(E - H_0 - V_\gamma)\psi_\beta^{(\gamma)} = V_\gamma \sum_\alpha \bar{\delta}_{\gamma\alpha}\psi_\beta^{(\alpha)}. \tag{53}$$

The integral form of Eq. (53) is

$$\psi_\beta^{(\gamma)} = \delta_{\gamma\beta}\phi_\beta + (E + i0 - H_0 - V_\gamma)^{-1} V_\gamma \sum_\alpha \bar{\delta}_{\gamma\alpha}\psi_\beta^{(\alpha)}. \tag{54}$$

The inhomogeneous term specifies the incoming boundary condition and is shown to have the above form in Appendix A where the limit $\epsilon \to 0$ is considered. The Eqs. (52) and (53) along with the specification of the components by Eq. (39) are the main results of this paper. In the case $N=3$ where V_γ is the interaction for the γ-pair, this formulation is identical to that of Faddeev.

Let us now consider whether Eq. (53) solves the problem of too many channel Hamiltonians for the N-body case as Eq. (10) does for the three-body case. The crucial property is the confinement of the source term in coordinate space. The irreducible potential V_γ consists only of graphs with γ-connectivity. Therefore, every term V_γ contains interactions of each particle within one of the clusters of γ with all of the other particles of that same cluster either directly or through a chain of interactions. Assuming the potentials to be short ranged, if any of the particles gets far away from its cluster, V_γ will vanish. The only coordinate which is not confined by V_γ is the relative distance between the center of mass of the two clusters. The support of V_γ can be therefore visualized as a tube in the many-body configuration space. We refer to this as the γ-tube.

Equation (52) says that $\psi_\beta^{(\gamma)}$ has its source within the γ-tube. Since the source term on the right of Eq. (53) contains the product of V_γ and $\psi_\beta^{(\alpha)}$ and $\alpha \neq \gamma$, it is described by the intersection of two distinct tubes in the N-body configuration space. As a result of this intersection, any pair of particles is connected by a string of interactions, and the source term in Eq. (53) is confined in the sense that it falls off faster than the kinetic energy term.

We note that as the number of particles is increased the confinement of the source may become weaker. The operator V_γ does not confine the clusters of γ to bound states. The outgoing wave of a breakup part of one of the clusters will only fall off with distance according to a power law. In the three-body problem this long-range effect is well known[24,25] and has important physical implications for both bound[26] and scattering states.[25] The long range character can be seen in the structure of V_γ in the four-body example given above. The presence of the Green function between the interactions in Eq. (51) and the dependence of the energy denominator of the two-body T matrix on the spectator momenta both lead to slowly decaying terms, even for potentials of short range. Since these effects have important physical consequences, their appearance here is necessary. If, in the treatment of a many-body problem, a trunca-

tion is made to eliminate breakup into more than m clusters, the long range effects associated with m and fewer cluster states should still be present.

Since the source term in Eq. (53) is confined, $\psi_\beta^{(\gamma)}$ satisfies

$$(E - H_0 - V_\gamma)\psi_\beta^{(\gamma)} = 0 \tag{55}$$

everywhere in the asymptotic region. The component $\psi_\beta^{(\gamma)}$ is therefore propagated asymptotically by the effective channel Hamiltonian

$$H_\gamma = H_0 + V_\gamma. \tag{56}$$

In the two-cluster part of the asymptotic region H_γ reduces to H_γ. To see this we apply the anticluster expansion, Eq. (43), to the definition of the irreducible potential, Eq. (48). This gives the anticluster expansion for V_γ,

$$V_\gamma = \sum_{(\gamma \supseteq)a} N(\gamma, a) V_a G_a G_\gamma^{-1}. \tag{57}$$

In the two-cluster region, the wavefunction will have the structure of the product of the two internal wavefunctions times a relative outgoing wave. Asymptotically the relative wavefunction must be on-shell so the entire function satisfies

$$G_\gamma^{-1}\phi_\gamma = 0. \tag{58}$$

The only term in $V_\gamma\phi_\gamma$ which can cancel the zero of the G_γ^{-1} is the term $\gamma = a$. We therefore have on the two-cluster asymptotic part

$$V_\gamma\phi_\gamma = V_\gamma\phi_\gamma. \tag{59}$$

The component wavefunction $\psi_\beta^{(\gamma)}$ also has outgoing waves corresponding to breakup. Those breakup waves found in this particular component correspond to bound clusters in partitions a ($\subset \gamma$). In this part of the asymptotic region, V_γ not only provides the internal interactions necessary to bind the clusters of a, but also that part of their interactions in which the last scattering is external to a but internal to γ. The full breakup wave for partition a is obtained by adding all those components $\psi_\beta^{(\gamma)}$ for which $\gamma \supset a$.

Finally we demonstrate that the on-shell transition amplitudes are precisely the coefficients of the outgoing waves of a component. This means that the Eqs. (53) could be solved in coordinate space and the scattering amplitude determined directly from the asymptotic form of the solution without the need for a quadrature. We show this here for the case of two-cluster final states. The demonstration of the similar result for breakup is somewhat more technical and is given in Appendix B.

From Eq. (39) the scattered wave in the components has the form

$$\psi_\beta^{(\gamma)s} = \psi_\beta^{(\gamma)} - \delta_{\beta\gamma}\phi_\beta = G_0 K_\gamma G_0 T^{\gamma\beta}\phi_\beta. \tag{60}$$

Applying the anticluster expansion to G_0K_γ [analogous to Eq. (43)] gives

$$G_0 K_\gamma = \sum_{(\gamma \supseteq)a} N(\gamma, a) G_a V_a. \tag{61}$$

Substituting this into Eq. (60) gives

$$\psi_\beta^{(\gamma)s} = \sum_{(\gamma \supseteq)a} N(\gamma, a) G_a V_a G_0 T^{\gamma\beta}\phi_\beta. \tag{62}$$

The only term in this sum which can have an outgoing two-cluster state is $a = \gamma$. If $\gamma \supset a$, then the operator $G_aV_aG_0$ cannot provide enough interactions to bind the particles into two separate clusters. The operator $T^{\gamma\beta}$ has a final V^γ [cf. Eq. (25)] which blocks the singularities of terms to its right and prevents them from building a two-cluster outgoing wave.[21]

Making a spectral expansion of G_γ gives

$$G_\gamma = \frac{1}{E + i\epsilon - T_\gamma - \epsilon_\gamma} |\phi_\gamma\rangle\langle\phi_\gamma| + \text{outgoing waves of three and more clusters.} \tag{63}$$

The operator T_γ is the relative kinetic energy of the two clusters and ϵ_γ is the sum of their internal energies. Applying the expansion (63) to Eq. (62) we have

$$\psi_\beta^{(\gamma)s} = \frac{1}{E + i\epsilon - T_\gamma - \epsilon_\gamma} |\phi_\gamma\rangle\langle\phi_\gamma| V_\gamma G_0 T^{\gamma\beta}|\phi_\beta\rangle + \text{outgoing waves of three and more clusters.} \tag{64}$$

The first term of Eq. (64) is the primary pole singularity which produces two-cluster outgoing waves.[17] Asymptotically only on-shell states ϕ_γ lead to nonvanishing contributions. Since on-shell

$$\langle\phi_\gamma| V_\gamma G_0 = \langle\phi_\gamma|, \tag{65}$$

the coefficient of the outgoing two-cluster state is

$$\langle\phi_\gamma| V_\gamma G_0 T^{\gamma\beta}|\phi_\beta\rangle = \langle\phi_\gamma| T^{\gamma\beta}|\phi_\beta\rangle, \tag{66}$$

the transition amplitude for the process.

VI. SUMMARY AND CONCLUSIONS

The fundamental formal problem of many-body scattering theory is the appearance of many-channel Hamiltonians needed to describe the asymptotic region. This difficulty is much more complex for N-bodies than for three because of the large number of possible arrangement and breakup channels. The three-body method of Faddeev overcomes this difficulty by obtaining an integral equation formulation in which the iterated kernel is connected. In terms of differential equations for the three wavefunction components, this latter condition translates into confinement of source terms. Thus the boundary conditions are well specified since each component is propagated in the entire asymptotic region by its own channel Hamiltonian.

We develop a wavefunction formulation for the N-body problem by making use of recent work on integral equations for transition operators. We translate the condition of a connected iterated kernel of the transition operator equations into a confinement property for the source terms of differential equations for components of the N-body wavefunction.

We find a division of the full wavefunction into two-cluster components which satisfy a set of coupled differential equations of Faddeev structure. These equations satisfy the basic requirements for a solution, namely

that there is a confined source term, and therefore that each component is propagated in the asymptotic region by its own effective channel Hamiltonian. Though our effective channel Hamiltonian is in general a complicated operator, in the region of asymptotic coordinate space relevant to the propagation of two-cluster states it reduces to the usual partition Hamiltonian. Each wavefunction component is labeled by a two-cluster partition index and has outgoing waves corresponding to bound and breakup states of that partition only. Scattering amplitudes can therefore be extracted directly from the coefficients of the outgoing waves of the relevant components.

In the present formulation we have chosen to label the components by two-cluster indices only and to construct components satisfying Faddeev-like equations. By "Faddeev-like equations" we mean equations having the algebraic structure of Eq. (10) in which the confinement of the sources arises from the property that, in them, any particle interacts with any other particle, either directly, or through a connected string of interactions involving other particles. (In the language of integral equations this is equivalent to saying that the first iterate of the kernel is completely connected.) As a result we are required to introduce operators of high connectivity. This necessary characteristic is satisfied by a two-cluster decomposition because of the unique property of two-cluster connected graphs: The conjoining of two subgraphs having two-cluster connectivity corresponding to any two distinct partitions always yields a completely connected graph.

A less highly connected operator could be employed, but as the degree of connectivity is lowered, additional iterations must be introduced before the kernel becomes connected. The two divisions of the N-body wavefunction into components which exist in the literature use kernels of low connectivity but employ different methods to guarantee confinement (connectivity). The division of Sandhas[27] introduces a chain of indices so that a wavefunction component is labeled by a sequence of partitions $a_{N-1} \subset a_{N-2} \cdots \subset a_3 \subset \gamma$. The effective potential is then only a single potential operator so it is as disconnected as possible. A string of restrictions on the coupling of successive indices produces confinement (connectivity) after N-2 iterations. The resulting equations have the form of the three-body Faddeev equation, (10), but all the operators are now matrices with indices being chains of partitions.

A second method is that of Kouri, Krüger, and Levin[28] who decompose as we do, by two-cluster indices, but retain a kernel of first order in the potential by foregoing the Faddeev structure. The components are coupled chainwise, each one only to a single other component. The connectivity arises from the fact that the product $V^{\alpha_1} G_0 V^{\alpha_2} \cdots G_0 V^{\alpha_n}$ contains only completely connected graphs if the partitions $\alpha_1, \ldots, \alpha_n$ include all the two-cluster partitions.

Each of the three methods has advantages and disadvantages. The division of Sandhas has the advantage that the kernel contains only simple operators (potentials) but has the disadvantage of introducing the most complete decomposition possible. This latter fact can be an advantage in the case $N=3$ or 4 but becomes a disaster for N significantly larger than that. The number of equations grows precipitously[29] with N and the partition indices not associated with channels of the asymptotic wavefunction suggest that the equations cannot be truncated in a simple manner.

The division of Kouri *et al.* has the double advantage of only having two-cluster indices and involving simple operators, but has the disadvantages of not always reducing to Faddeev equations for the treatment of three-body degrees of freedom and of depending on a choice of ordering the two-cluster partitions. Their equations do not explicitly preserve the symmetry group of exchange of partition labels. This causes no trouble in an exact treatment, but in any truncation different choices may lead to different results. The implication of the symmetry breaking is not yet fully understood.

In our method, the index structure is simple and the connectivity obtained very directly. Many of the difficulties of the N-body problem are suppressed into understanding the structure of the complicated operators, V_γ. Yet the explicit forms we have for these operators, Eqs. (49) and (57), show that the sybsystem Green functions appear inside V_γ in straightforward ways. The highly clustered states which dominate most reactions can thereby be easily introduced. One may therefore have the hope of finding truncations of our equations which form a bridge between the standard two-cluster methods of reaction theory[30] and the recently developed methods of the three-body problem.[31] In addition, we hope that by giving us new insights into the structure of the N-body scattering wavefunction these results should lead to the development of more versatile approximate methods and a firmer basis for a theory of reactions than we now possess.

ACKNOWLEDGMENTS

We would like to thank Frank Levin for his encouraging comments. Two of us (M. L. H. and P. C. T.) acknowledge the support of the University of Maryland where much of this work was done. Two of us (E. F. R. and P. C. T.) are grateful to the IPN, Orsay for their hospitality.

APPENDIX A

In this appendix we demonstrate that the limit $\epsilon \to 0$ is properly handled in the derivation of the final result, Eqs. (53) and (54). The basic difficulty is that the Lippmann identity, Eq. (45), does not necessarily hold when multiplied from the left by an operator before the ϵ limit is taken. It holds only under certain restrictive conditions. Essentially, these conditions state that each term in the perturbation expansion characterizing the multiplicative operator, when followed from the right, must have an interaction residual to β before it has an infinite number of each of the interactions internal to β occuring in all possible sequences.[19,22] If the condition were not satisfied, a Green function G_β would be built and the operator would contain a string of the form $G_\beta V_\beta$ leading to $(\alpha \neq \beta)$

$$G_\beta V_\beta (G_\alpha G_\beta^{-1} \phi_\beta)$$

$$= [G_\beta - G_\alpha + G_\beta V_\alpha G_\alpha] G_\beta^{-1}\phi_\beta$$
$$= [G_\alpha (V_\beta - V_\alpha) G_\beta + G_\beta V_\alpha G_\alpha] G_\beta^{-1}\phi_\beta$$
$$= G_\alpha (V_\beta - V_\alpha)\phi_\beta + G_\beta V_\alpha G_\alpha G_\beta^{-1}\phi_\beta, \tag{A1}$$

which does not vanish in the limit $\epsilon \to 0$ as would be required by Eq. (45). To get (A1) we have used

$$G_\beta = G_\alpha + G_\alpha (V_\beta - V_\alpha) G_\beta$$
$$= G_\alpha + G_\beta (V_\beta - V_\alpha) G_\alpha. \tag{A2}$$

In Eqs. (53) and (54) the wavefunctions and hence the Lippmann identity are multiplied by operators from the left. What should be done is to take none of the limits $\epsilon \to 0$ until after all manipulations have been carried out. This means that instead of Eq. (29) we should employ the version of the BRS equations in which the inhomogeneous term is not transformed using the limit $\epsilon \to 0$.

Sketching this briefly, we begin with Eqs. (23) and (25) as definitions of ψ_β and $T^{\alpha\beta}$ for $\epsilon \neq 0$. Instead of Eq. (28) we then get

$$\psi_\beta = G_0 V \psi_\beta + G_0 G_\beta^{-1}\phi_\beta. \tag{A3}$$

Equation (26) is still correct, so with (A3) we obtain

$$\psi_\beta = G_0 T^{0\beta}\phi_\beta + G_0 G_\beta^{-1}\phi_\beta, \tag{A4}$$

instead of Eq. (27).

The untransformed BRS equations for $T^{0\beta}$ take the form[11]

$$T^{0\beta}\phi_\beta = \sum_c K_c G_0 G_\beta^{-1}\phi_\beta + \sum_\gamma K_\gamma G_0 T^{\gamma\beta}\phi_\beta. \tag{A5}$$

Putting this into Eq. (A4) gives

$$\psi_\beta = G_0 G_\beta^{-1}\phi_\beta + \sum_c G_0 K_c G_0 G_\beta^{-1}\phi_\beta + \sum_\gamma G_0 K_\gamma G_0 T^{\gamma\beta}\phi_\beta. \tag{A6}$$

Introducing components defined as before by Eq. (39), we now get

$$\psi_\beta = \sum_\gamma \psi_\beta^{(\gamma)} + [G_0 - G_\beta + \sum_c G_0 K_c G_0] G_\beta^{-1}\phi_\beta, \tag{A7}$$

instead of Eq. (38). To get a system of equations for the components we now express $T^{\gamma\beta}\phi_\beta$ in terms of ψ_β by using Eq. (41). Substitution of Eq. (41) into Eq. (39) still yields Eq. (42) for finite ϵ. However, since the division of ψ_β into components is now given by Eq. (A7) instead of Eq. (38), we obtain the system of equations

$$(E + i\epsilon - H_0 - V_\gamma)\psi_\beta^{(\gamma)} = V_\gamma \sum_\alpha \bar{\delta}_{\gamma\alpha}\psi_\beta^{(\alpha)} + (\delta_{\gamma\beta} G_0^{-1} - V_\gamma G_\gamma G_\beta^{-1})\phi_\beta + V_\gamma [G_0 - G_\beta + \sum_c G_0 K_c G_0] G_\beta^{-1}\phi_\beta. \tag{A8}$$

The integral form of these equations is obtained by multiplying both sides by $\mathcal{G}_\gamma$ defined as

$$\mathcal{G}_\gamma = (E + i\epsilon - H_0 - V_\gamma)^{-1}. \tag{A9}$$

The contribution from the second term on the right of Eq. (A8) is

$$\mathcal{G}_\gamma (\delta_{\gamma\beta} G_0^{-1} - V_\gamma G_\gamma G_\beta^{-1})\phi_\beta = \delta_{\gamma\beta}\mathcal{G}_\gamma (G_0^{-1} - V_\gamma)\phi_\beta - \bar{\delta}_{\gamma\beta}\mathcal{G}_\gamma V_\gamma G_\gamma G_\beta^{-1}\phi_\beta = \delta_{\gamma\beta}\phi_\beta - \bar{\delta}_{\gamma\beta}\mathcal{G}_\gamma V_\gamma G_\gamma G_\beta^{-1}\phi_\beta. \tag{A10}$$

We now take the limit $\epsilon \to 0$. Since $\mathcal{G}_\gamma V_\gamma$ is γ-connected, it must be missing at least one interaction internal to β, and therefore cannot provide a singularity to cancel the zero in $G_\beta^{-1}\phi_\beta$. So Eq. (45) may be used and the standard incoming boundary condition $\delta_{\gamma\beta}\phi_\beta$ is obtained. The vanishing of the second term of Eq. (A10) is still valid when the factor $\mathcal{G}_\gamma$ is removed, since V_γ is still γ-connected. So in the differential equation form, the contribution is just $\delta_{\gamma\beta}\mathcal{G}_\gamma^{-1}\phi_\beta$, which, because of Eqs. (58) and (20), vanishes in the limit.

To treat the third term on the right of Eq. (A8) we use the cluster expansion of $G_\beta - G_0$. This takes the form[11]

$$G_\beta - G_0 = G_0 V_\beta G_\beta = \sum_{\beta \supseteq c} G_0 K_c G_0. \tag{A11}$$

Apart from the simple factor G_0, the inverse of this expansion is what we have given in Eq. (43). Essentially, Eq. (A11) is just a classification of the perturbation graphs of $G_\beta - G_0$ into Weinberg graphs of connectivities equal to or contained in β. The third term on the right of Eq. (A8) becomes

$$V_\gamma \sum_{\beta \not\supseteq c} G_0 K_c G_0 G_\beta^{-1}\phi_\beta. \tag{A12}$$

We now take the limit $\epsilon \to 0$. Because of the condition $\beta \not\supseteq c$, every term K_c that enters has at least one interaction external to β, and is missing at least one of the interactions internal to β. This fulfils the condition stated above and Eq. (45) can be used to show the vanishing of this term. Multiplication by $\mathcal{G}_\gamma$ from the left prior to the limit leaves the above reasoning unchanged and this term does not contribute to the integral or differential forms of Eq. (A8). This verifies that Eqs. (53) and (54) hold in the limit $\epsilon \to 0$.

APPENDIX B

We outline here the structure of our wavefunction components for outgoing breakup waves (i.e., for final states of more than two clusters). In particular we demonstrate that the coefficient of an outgoing breakup wave is the corresponding on-shell transition amplitude. This case is more difficult than the two-cluster states treated in Sec. V because our wavefunction components are labeled only by two-cluster partitions. Thus in the anticluster expansion for $G_0 K_\gamma$ [see Eq. (61)] there will be more than one term contributing to a given multicluster outgoing wave.

We overcome this difficulty by revealing the subpartition structure of K_γ. Equations (48) and (49) combine to give

$$K_\gamma G_0 = \sum_{a_{N-1} \subset a_{N-2} \subset \cdots \subset a_3 (\subset \gamma)} V_{a_{N-1}} G_{a_{N-1}} \times V_{a_{N-2}}^{a_{N-1}} G_{a_{N-2}} \cdots G_{a_3} V_\gamma^{a_3} G_\gamma. \tag{B1}$$

If we omit all terms to the right of G_{a_n}, we have left

the sum of all Weinberg graphs of connectivity a_n. This is just the definition of K_{a_n}. Explicitly

$$K_{a_n}G_0 = \sum_{a_{N-1}\subset a_{N-2}\subset\cdots\subset a_{n+1}\subset a_n} V_{a_{N-1}}G_{a_{N-1}}V^{a_{N-1}}_{a_{N-2}}G_{a_{N-2}} \cdots G_{a_{n+1}}V^{a_{n+1}}_{a_n}G_{a_n}. \tag{B2}$$

Combining Eqs. (B2) and (B1) we have

$$K_\gamma G_0 = \sum_{a_n\subset a_{n+1}\subset\cdots\subset a_3\subset\gamma} K_{a_n}G_0V^{a_n}_{a_{n-1}}G_{a_{n-1}}\cdots G_{a_3}V^{a_3}_\gamma G_\gamma. \tag{B3}$$

Substituting Eq. (B3) into Eq. (60), the total scattered wave can be written

$$\sum_\gamma \psi^{(\gamma)s}_\beta = G_0\sum_\gamma K_\gamma G_0 T^{\gamma\beta}\phi_\beta = G_0\sum_{a_n} K_{a_n}G_0\tilde{T}^{a_n\beta}\phi_\beta, \tag{B4}$$

where n is fixed, and

$$\tilde{T}^{a_n\beta} = \sum_{(a_n\subset)a_{n+1}\subset\cdots a_3\subset\gamma} V^{a_n}_{a_{n-1}}G_{a_{n-1}}\cdots G_{a_3}V^{a_3}_\gamma G_\gamma T^{\gamma\beta}. \tag{B5}$$

With Eq. (B4) we are now in a position to look at outgoing waves of n bound clusters. Let the corresponding partition be a, where $n_a = n$. Applying the anticluster expansion [analogous to Eq. (61)]

$$G_0K_a = \sum_{(a\subset)b} N(a,b)G_bV_b \tag{B6}$$

to the appropriate term of Eq. (B4) gives

$$G_0K_aG_0\tilde{T}^{a\beta}\phi_\beta = \sum_{(a\subset)b} N(a,b)G_bV_bG_0\tilde{T}^{a\beta}\phi_\beta. \tag{B7}$$

The only term that can have an outgoing wave of the n clusters specified by partition a is $b=a$. The method now parallels exactly that used for Eq. (62). The result is that the coefficient of the outgoing wave is the on-shell matrix element

$$\langle\phi_a|\tilde{T}^{a\beta}|\phi_\beta\rangle. \tag{B8}$$

We now establish that this on-shell matrix element is identical to the on-shell transition amplitude $\langle\phi_a|T^{a\beta}|\phi_\beta\rangle$. Consider a wavefunction defined according to

$$\psi'_\beta = G_0\sum_{a_n} K_{a_n}G_0T^{a_n\beta}\phi_\beta, = G_0\left(\sum_{a_n}K_{a_n}G_0T^{a_n0}\right)G_0G_\beta^{-1}\phi_\beta, \tag{B9}$$

where Eqs. (21) and (25) have been used. The summation extends over partitions having n clusters. From the definition of K_{a_n} and the structure [Eq. (24)] of T^{a_n0}, the term in square brackets can be characterized as the sum of all Weinberg graphs having connectivity greater than that of an n-cluster graph. We can divide it into two parts:

(1) the sum of all fully connected Weinberg graphs, plus

(2) the sum of all Weinberg graphs having connectivity greater than that of an n-cluster graph but less than or equal to the connectivity of a two-cluster graph.

Using the operator description of each of these two parts, Eq. (B9) becomes

$$\psi'_\beta = G_0\left[\sum_\gamma K_\gamma G_0T^{\gamma0} + \sum_{m=2}^{n-1}\sum_{a_m}K_{a_m}\right]G_0G_\beta^{-1}\phi_\beta = G_0\sum_\gamma K_\gamma G_0T^{\gamma\beta}\phi_\beta + G_0\sum_{m=2}^{n-1}\sum_{a_m}K_{a_m}G_0G_\beta^{-1}\phi_\beta, \tag{B10}$$

where Eqs. (21) and (25) have been used on the first term.

The second term simplifies in the limit $\epsilon\to 0$. Only the term $a_m=\beta$ provides enough interactions to cancel the zero in $G_\beta^{-1}\phi_\beta$, and by Eqs. (46) and (20), we have

$$\psi'_\beta = G_0\sum_\gamma K_\gamma G_0T^{\gamma\beta}\phi_\beta + \phi_\beta. \tag{B11}$$

Therefore, by Eq. (30), ψ'_β equals the full wavefunction ψ_β. Combining Eqs. (B4), and (B9) and (B11) yields

$$G_0\sum_{a_n}K_{a_n}G_0(T^{a_n\beta}-\tilde{T}^{a_n\beta})\phi_\beta = \phi_\beta. \tag{B12}$$

Thus for any breakup channel corresponding to a partition a_n the coefficients of the outgoing waves in Eq. (B12) are all zero, and hence the on-shell matrix elements of $T^{a_n\beta}$ and $\tilde{T}^{a_n\beta}$ are identical. To be more explicit, applying the anticluster expansion [Eq. (B6)] for K_{a_n} in Eq. (B12), and using $\langle\phi_{a_n}|V_{a_n}G_0 = \langle\phi_{a_n}|$ for on-shell states, we obtain, by the method analogous to that employed for Eq. (62), the result

$$\langle\phi_{a_n}|\tilde{T}^{a_n\beta}|\phi_\beta\rangle = \langle\phi_{a_n}|T^{a_n\beta}|\phi_\beta\rangle \tag{B13}$$

on the energy shell.

[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 the Quantum Scattering Theory* (Israel Program for Scientific Translation, Jerusalem, 1965).
[3]E. Schmid and H. Ziegelmann, *The Quantum Mechanical Three-Body Problem*, Vieweg Tracts in Pure and Applied Physics, Vol. 2 (Vieweg, Braunschweig, 1974).
[4]S. Weinberg, Phys. Rev. **133**, B232 (1964).
[5]L. Rosenberg, Phys. Rev. **140**, B217 (1965).
[6]O.A. Yakubovskii, Yad. Fiz. **5**, 1312 (1967) [Sov. J. Nucl. Phys. **5**, 937 (1967)].
[7]P. Grassberger and W. Sandhas, Nucl. Phys. B **2**, 181 (1967).
[8]R. Omnes, Phys. Rev. **165**, 1265 (1968).
[9]I.H. Sloan, Phys. Rev. C **6**, 1945 (1972).
[10]Gy. Bencze, Nucl. Phys. A **210**, 568 (1973).
[11]E.F. Redish, Nucl. Phys. A **225**, 16 (1974). Note that in this reference the operator denoted by W_a^γ is identical to our present K_γ^a.
[12]D.J. Kouri and F.S. Levin, Phys. Lett. B **50**, 421 (1974).
[13]W. Tobocman, Phys. Rev. C **9**, 2466 (1974).
[14]D.J. Kouri and F.S. Levin, Nucl. Phys. A **253**, 395 (1975).
[15]V. Vanzani, Lett. Nuovo Cimento **16**, 1 (1976).
[16]W. Glöckle, Nucl. Phys. A **141**, 620 (1970).
[17]T.A. Osborn and D. Bollé, Phys. Rev. C **8**, 1198 (1973).
[18]M. Goldberger and K.M. Watson, *Collision Theory* (Wiley, New York, 1964).

[19]P. Benoist-Gueutal, M. L'Huillier, E.F. Redish, and P.C. Tandy (to be published in Phys. Rev. C).
[20]E.F. Redish, Nucl. Phys. A **235**, 82 (1974).
[21]E.F. Redish, P.C. Tandy, and M. L'Huillier, Phys. Lett. B **61**, 413 (1976).
[22]R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).
[23]B.A. Lippmann, Phys. Rev. **102**, 264 (1956).
[24]H.P. Noyes and H. Fiedeldey, in *Three Particle Scattering in Quantum Mechanics*, edited by J. Gillespie and J. Nuttall (Benjamin, New York, 1968), p. 195.
[25]H.P. Noyes, Phys. Rev. Lett. **23**, 1201 (1969).
[26]V. Efimov, in *Few Body Dynamics*, edited by A. Mitra *et al.* (North-Holland, Amsterdam, 1976), p. 126.
[27]W. Sandhas, in *Few Body Dynamics*, edited by A. Mitra *et al.* (North-Holland, Amsterdam, 1976), p. 540.
[28]D.J. Kouri, H. Krüger, and F.S. Levin, Phys. Rev. D **15**, 1156 (1977).
[29]Gy. Bencze, "Combinatorial Problems in N-Particle Scattering Theories," International Symposium on Nuclear Reaction Models, Balatonfüred, Hungary, June 27—July 1, 1977 (to be published); Phys. Lett. B **72**, 155 (1977).
[30]K. Wildermuth and Y.C. Tang, *A Unified Theory of the Nucleus* (Vieweg, Braunschweig, 1977); N.F. Mott and H.S.W. Massey, *The Theory of Atomic Collisions*, 3rd edition (Oxford U.P., Oxford, 1965).
[31]S.P. Merkuriev, C. Gignoux, and A. Laverne, Ann. Phys. **99**, 30 (1976).

A class of multidimensional nonlinear Langmuir waves

P. K. C. Wang

School of Engineering and Applied Science, University of California, Los Angeles, California 90024
(Received 3 October 1977)

Nonlinear Langmuir waves in a plasma governed by the dimensionless equations $i\partial \mathbf{E}/\partial t = -\nabla^2\mathbf{E} + n\mathbf{E}$, $\partial^2 n/\partial t^2 = \nabla^2[n + g(|\mathbf{E}|^2)]$ are studied, where $\mathbf{E}$ is the complex amplitude of the high-frequency electric field, n is the low frequency perturbation in the ion density from its constant equilibrium value, and g is a given function of $|\mathbf{E}|^2$. General conditions for the existence or nonexistence of a class of multidimensional solitary-wave and nonlinear periodic travelling-wave solutions in the form $\mathbf{E}(t,\mathbf{x}) = \mathbf{h}(\mathbf{k}\cdot\mathbf{x} - vt)$ and $n(t,\mathbf{x}) = s(\mathbf{k}\cdot\mathbf{x} - vt)$ are established. The results are applied to the special cases: (i) $g(|\mathbf{E}|^2) = |\mathbf{E}|^2$ corresponding to the usual pondermotive force, and (ii) $g(|\mathbf{E}|^2) = K[1 - \exp(-|\mathbf{E}|^2)]$, K is a positive constant, representing ion density saturation.

1. INTRODUCTION

The formation, interaction and collapse of nonlinear Langmuir waves in plasmas have been studied extensively in recent years.[1-10] In most of the existing works, attention is focused primarily on the formation and interaction of solitary waves. Exact expressions for these solitary waves for various regimes have been obtained only for the one-dimensional case. Recently, Gibbons *et al.*[10] discussed the possibility of existence of solitary Langmuir waves for higher dimensions. In this study, we obtain conditions for the existence or nonexistence of a class of multidimensional nonlinear Langmuir travelling waves including the periodic and the usual one-dimensional solitary waves.

We begin with the following basic equations[1] describing the nonlinear interaction of high-frequency electron oscillations with an ion fluid:

$$\begin{aligned} &i\partial\mathbf{E}/\partial t = -\nabla^2\mathbf{E} + n\mathbf{E}, \\ &\partial^2 n/\partial t^2 = \nabla^2[n + g(|\mathbf{E}|^2)], \end{aligned} \tag{1}$$

where $i = \sqrt{-1}$, $\mathbf{E} = (E_1, \ldots, E_N)$ is the complex amplitude of the high-frequency electric field $\mathcal{E}$ given by

$$\mathcal{E}(t,\mathbf{x}) = \mathrm{Re}[\mathbf{E}(t,\mathbf{x})\exp(-i\omega_p t)]; \tag{2}$$

n is a real quantity corresponding to the low-frequency perturbation in the ion density from its constant equilibrium value n_0; and g is a specified real-valued function of $|\mathbf{E}|^2$. Here, we have used dimensionless quantities. The units of time t, spatial coordinates $\mathbf{x} = (x_1, \ldots, x_N)$, electric field $\mathbf{E}$ and density n are, respectively, $3q/(2\alpha\omega_p)$, $(3r_D/2)(q\alpha)^{-1/2}$, $8(q\alpha n_0\pi T/3)$, and $4q\alpha n_0/3$, where α is the electron–ion mass ratio m_e/m_i, $q = T/T_e$, $T = T_e + T_i$, T_e, T_i the electron and ion temperatures respectively, r_D the electron Debye radius, and ω_p is the plasma frequency. The function g is introduced here so as to permit the consideration of a wide class of nonlinear effects such as saturation.

2. TRAVELLING WAVE SOLUTIONS

Let $\mathbb{R}^N$ and $\mathbb{C}^N$ denote the N-dimensional real and complex Euclidean spaces respectively, and $C_m(\mathbb{R}; V)$ the space of all m-times continuously differentiable functions defined on $\mathbb{R}$ and taking their values in the vector space V. The norms for $\mathbb{R}^N$ and $\mathbb{C}^N$ are denoted by $\|\cdot\|$ and $|\cdot|$ respectively. The dot notation is used to denote the usual scalar product on $\mathbb{R}^N$ or $\mathbb{C}^N$.

Let $\mathbf{k}$ be a specified unit vector in $\mathbb{R}^N$ and v a given real number corresponding to a constant dimensionless velocity. We seek travelling-wave solutions of (1) in the form

$$\begin{aligned} &\mathbf{E}(t,\mathbf{x}) = \mathbf{h}(\mathbf{k}\cdot\mathbf{x} - vt), \\ &n(t,\mathbf{x}) = s(\mathbf{k}\cdot\mathbf{x} - vt), \end{aligned} \tag{3}$$

where $\mathbf{h}$ and s are undetermined functions in $C_2(\mathbb{R}, \mathbb{C}^N)$ and $C_2(\mathbb{R}, \mathbb{R})$ respectively. For physical reasons, we shall restrict $\mathbf{h}$ and s to functions such that $|\mathbf{h}(\xi)|$ and $|s(\xi)|$ are uniformly bounded on $\mathbb{R}$, where $\xi = \mathbf{k}\cdot\mathbf{x} - vt$. In particular, we shall consider multidimensional solitary-wave solutions which are analogous to those in the one-dimensional case. Here, we require that $|\mathbf{h}(\xi)|$ and $s(\xi)$ tend to finite values as $|\xi| \to \infty$.

Substituting (3) into (1) leads directly to the following equations for $\mathbf{h} = (h_1, \ldots, h_N)$ and s:

$$-iv\,d\mathbf{h}/d\xi + d^2\mathbf{h}/d\xi^2 = s(\xi)\mathbf{h}(\xi), \tag{4}$$

$$(v^2 - 1)d^2 s/d\xi^2 = d^2 g(|\mathbf{h}|^2)/d\xi^2, \tag{5}$$

where we have adopted the rectangular Cartesian coordinate system.

Equation (5) can be integrated to give

$$(v^2 - 1)s(\xi) = g(|\mathbf{h}(\xi)|^2) + \hat{C}\xi + C, \tag{6}$$

where $\hat{C}$ and C are integration constants. From the boundedness requirement, we set $\hat{C} = 0$. Assuming that $v^2 \neq 1$, we can solve for $s(\xi)$ in (6) and substitute it into (4) to give a complex differential equation for $\mathbf{h}$:

$$d^2\mathbf{h}/d\xi^2 - iv\,d\mathbf{h}/d\xi = (v^2 - 1)^{-1}[g(|\mathbf{h}(\xi)|^2) + C]\mathbf{h}. \tag{7}$$

It is advantageous to rewrite (7) in polar form. Let $h_j(\xi) = A_j(\xi)\exp[i\theta_j(\xi)]$, $j = 1, \ldots, N$. Then, we have

$$d^2A_j/d\xi^2 + A_j\theta_j'(\xi)[v - \theta_j'(\xi)] = (v^2 - 1)^{-1}A_j[g(\|\mathbf{A}\|^2) + C], \tag{8}$$

$$d^2\theta_j/d\xi^2 = [v - 2\theta_j'(\xi)]d(\ln A_j)/d\xi, \quad j = 1, \ldots, N, \tag{9}$$

where $\mathbf{A} = (A_1, \ldots, A_N)$, $\|\mathbf{A}\| = |\mathbf{h}|$ and $\theta_j' = d\theta_j/d\xi$.

Equation (9) can be integrated to give

$$\theta_j'(\xi) = [v - \mu_j A_j^{-2}(\xi)]/2, \tag{10}$$

where $\mu_j = A_j^2(0)[v - 2\theta_j'(0)]$.

Substituting (10) into (8) leads to the following differential equations for A_j:

$$d^2A_j/d\xi^2 = f(\mu_j, C, \mathbf{A})A_j, \quad j = 1, \ldots, N, \tag{11}$$

 0022-2488/78/1906-1286$1.00

where

$$f(\mu_j, C, \mathbf{A}) \triangleq (\mu_j^2 A_j^{-4} - v^2)/4 + (v^2-1)^{-1}[g(\|\mathbf{A}\|^2) + C]. \quad (12)$$

Evidently, given C, $\theta_j'(0)$, $A_j(0)$, $A_j'(0)$, $j=1,\ldots,N$, (11) can be integrated independently. Since f is a function tion of C and μ_j [depending on $A_j(0)$], (11) must be solved with initial conditions at $\xi=0$ which are consistent with the $A_j(0)$ in μ_j. Also, only those portions of solutions of (11) with $\mathbf{A}(\xi) \geq 0$ [i.e., $A_j(\xi) \geq 0$, $j=1,\ldots,N$] are meaningful here.

We note that (11) can be rewritten in the form:

$$d^2A_j/d\xi^2 = \partial U/\partial A_j, \quad j=1,\ldots,N, \quad (13)$$

where

$$U(\mathbf{A}, \mu, C) \triangleq U_1(\|\mathbf{A}\|^2, C) - \sum_{j=1}^{N} \mu_j^2/(8A_j^2), \quad (14)$$

$$2U_1(\|\mathbf{A}\|^2, C) \triangleq \int_0^{\|\mathbf{A}\|^2} [\kappa g(\eta) + \gamma]\, d\eta, \quad (15)$$

$$\kappa = (v^2-1)^{-1}, \quad \gamma = (v^2-1)^{-1}C - v^2/4, \quad (16)$$

and $\mu = (\mu_1, \ldots, \mu_N)$. A first integral of (13) is given by

$$I(\mathbf{A}(\xi), \mathbf{A}'(\xi)) \triangleq \|\mathbf{A}'(\xi)\|^2 - 2U_1(\|\mathbf{A}(\xi)\|^2, C) - \sum_{j=1}^{N} \mu_j^2 A_j^{-2}(\xi)/4 = C_1, \quad (17)$$

where $\|\mathbf{A}'(\xi)\|^2 \triangleq \sum_{j=1}^{N} [dA_j(\xi)/d\xi]^2$ and

$$C_1 = \|\mathbf{A}'(0)\|^2 - 2U_1(\|\mathbf{A}(0)\|^2, C) - \sum_{j=1}^{N} [v - 2\theta_j'(0)]. \quad (18)$$

Evidently, if $\mu_j \neq 0$ for some j, then $I(\mathbf{A}(\xi), \mathbf{A}'(\xi)) \to -\infty$ as $\|\mathbf{A}'(\xi)\|$ and $\|\mathbf{A}(\xi)\| \to 0$. Since C_1 is finite for finite $\|\mathbf{A}'(0)\|$, $\|\mathbf{A}(0)\|$, C, and $\theta_j'(0)$, $j=1,\ldots,N$, therefore there do not exist solutions of (13) or solitary-wave solutions of (7) such that $\|\mathbf{A}(\xi)\|$ and $\|\mathbf{A}'(\xi)\| \to 0$ as $|\xi| \to \infty$ when $\mu_j \neq 0$ for some j.

In what follows, we shall focus attention on the particular case where $\mu = 0$. Here, we have

$$\theta_j(\xi) = \theta_j(0) + v\xi/2, \quad j=1,\ldots,N \quad (19)$$

as a solution of (10) or (9). Note that $\mu_j = 0$ when $A_j(0) = 0$ and/or $\theta_j'(0) = v/2$. This implies that along any trajectory of (8), (9) starting from a point $z(0) = (\mathbf{A}(0), \mathbf{A}'(0), \theta(0), \theta'(0))$ in the set $Z \triangleq \{(\mathbf{A}, \mathbf{A}', \theta, \theta') \in \mathbb{R}^{4N}: A_j(v - 2\theta_j') = 0, j=1,\ldots,N\}$, its corresponding phase $\theta(\xi) \triangleq (\theta_1(\xi), \ldots, \theta_N(\xi))$ has the form (19). In this case, f no longer depends on $\mathbf{A}(0)$ and $\theta'(0)$, and (13) reduces to

$$d^2A_j/d\xi^2 = \partial U_1/\partial A_j, \quad j=1,\ldots,N. \quad (20)$$

The equilibrium points of (20) are points $(\mathbf{A}_e, 0)$ in $\mathbb{R}^{2N}$ such that $\mathbf{A}_e$ are the stationary points of U_1 or the roots of the equation $f(0, C, \mathbf{A})\mathbf{A} = 0$. Obviously, $\mathbf{A}_e$'s include $\mathbf{A}=0$ and all those $\mathbf{A}$'s satisfying $g(\|\mathbf{A}\|^2) = v^2(v^2-1)/4 - C$.

To obtain some qualitative information on the solutions of (20), we first derive a differential equation for $u(\xi) \triangleq \|\mathbf{A}(\xi)\|^2$. By direct computation,

$$d^2u/d\xi^2 = 2\|\mathbf{A}'(\xi)\|^2 + 2\mathbf{A}(\xi)\cdot d^2\mathbf{A}/d\xi^2 = 2\|\mathbf{A}'(\xi)\|^2 + 2u\tilde{f}(C, u), \quad (21)$$

where $\tilde{f}(C, \|\mathbf{A}\|^2) \triangleq f(0, C, \mathbf{A})$ as defined by (12). Along an integral curve (17) corresponding to a fixed C_1 and $\mu = 0$, (21) can be rewritten as

$$d^2u/d\xi^2 = 2[u\tilde{f}(C, u) + C_1 + 2U_1(u, C)] \triangleq P(u, C, C_1). \quad (22)$$

Its solution, starting with initial conditions

$$u(0) = \|\mathbf{A}(0)\|^2, \quad u'(0) = 2\mathbf{A}(0)\cdot\mathbf{A}'(0), \quad (23)$$

satisfying

$$\|\mathbf{A}'(0)\|^2 = C_1 + 2U_1(\|\mathbf{A}(0)\|^2, C) \geq 0, \quad (24)$$

describes the evolution of $\|\mathbf{A}(\xi)\|$ with ξ along the integral curve.

A first integral of (22) is given by

$$[u'(\xi)]^2 = [u'(0)]^2 + \int_{u(0)}^{u(\xi)} 2P(\eta, C, C_1)\, d\eta \triangleq Q(u, C, C_1, u'(0)), \quad (25)$$

where $u' = du/d\xi$. Equation (25) is valid only when its right-hand side is nonnegative. An implicit expression for $\|\mathbf{A}(\xi)\|^2$ can be obtained by integrating (25).

$$\int_{\|\mathbf{A}(0)\|^2}^{\|\mathbf{A}(\xi)\|^2} Q(\eta, C, C_1, u'(0))^{-1/2}\, d\eta = \pm\xi. \quad (26)$$

Note that if an explicit expression for $\|\mathbf{A}(\xi)\|^2$ is obtainable from (26), then $\mathbf{A}(\xi)$ can be determined by integrating each equation in (11) independently with $\mu = 0$.

In the sequel, we shall establish conditions for the existence or nonexistence of solutions of (20) having the property that $\|\mathbf{A}(\xi)\| \to 0$ as $|\xi| \to \infty$, or solitary-wave solutions of (7) with $\mu = 0$.

Theorem 1: If

$$\kappa g(u) + \gamma \geq 0 \quad (27)$$

for all $u \geq 0$, then there do not exist solutions of (20) such that $\|\mathbf{A}(0)\| > 0$ and $\|\mathbf{A}(\xi)\| \to 0$ as $|\xi| \to \infty$.

Proof: Condition (27) is equivalent to $\tilde{f}(C, u) \geq 0$ for all $u \geq 0$. In view of (21), we have $d^2u/d\xi^2 \geq 0$ implying that $\|\mathbf{A}(\xi)\|^2$ along any solution of (20) is a convex function of ξ. Hence, it is impossible to have $\|\mathbf{A}(0)\| > 0$ and $\|\mathbf{A}(\xi)\| \to 0$ as $|\xi| \to \infty$. ■

Note that for the subsonic ($v^2 < 1$) and supersonic ($v^2 > 1$) cases, (27) implies that $g(u)$ is uniformly bounded above and below by $v^2(v^2-1)/4 - C$. Also, if (27) is a strict inequality, then $(\mathbf{A}, \mathbf{A}') = (0, 0)$ is the only equilibrium point of (20).

Theorem 2: Assume that the following conditions are satisfied:

(i) $v^2(v^2-1) > 4C$ and $v^2 < 1$;

(ii) g is a strictly monotone increasing function in $C_1(\mathbb{R}, \mathbb{R})$ with $g(0) = 0$, and there exists a positive number $u_1 < \infty$ such that

$$\int_0^{u_1} g(\eta)\, d\eta = [v^2(v^2-1)/4 - C]u_1 \quad (28)$$

and

$$\int_0^{u} g(\eta)\, d\eta > [v^2(v^2-1)/4 - C]u \quad \text{for all } u > u_1. \quad (29)$$

Then (20) has a solution $\mathbf{A}(\xi) \geq 0$ for all $\xi \in \mathbb{R}$, with $\|\mathbf{A}(0)\| > 0$ and $\|\mathbf{A}'(0)\| = 0$ such that

$$\|\mathbf{A}(\xi)\| \text{ and } \|\mathbf{A}'(\xi)\| \to 0 \text{ as } |\xi| \to \infty. \tag{30}$$

Proof: First, we note from (15) and (17) with $\mu = 0$ that for a solution to have property (30), the initial condition $(\mathbf{A}(0), \mathbf{A}'(0))$ must satisfy

$$\tilde{C}_1 \triangleq \|\mathbf{A}'(0)\|^2 - 2U_1(\|\mathbf{A}(0)\|^2, C) = 0. \tag{31}$$

We shall show that, under condition (i), $\tilde{C}_1 = 0$ implies property (30). From (17), it is evident that when $(\mathbf{A}(0), \mathbf{A}'(0))$ satisfies (31), its corresponding trajectory is a zero-level curve of $I(\mathbf{A}, \mathbf{A}')$ defined by

$$I(\mathbf{A}, \mathbf{A}') \triangleq \|\mathbf{A}'\|^2 - 2U_1(\|\mathbf{A}\|^2, C) = 0, \tag{32}$$

or the points along the trajectory belong to the set

$$\Gamma^1(0) = \{(\mathbf{A}, \mathbf{A}') \in \mathbb{R}^{2N} : \|\mathbf{A}'\|^2 = 2U_1(\|\mathbf{A}\|^2, C)\}. \tag{33}$$

Obviously, the equilibrium point $(\mathbf{A}, \mathbf{A}') = (0, 0) \in \Gamma^1(0)$. Now the foregoing implications can be established by verifying that $(0,0)$ is the only equilibrium point in $\Gamma^1(0)$; moreover, it is a saddle point.

Let $(\mathbf{A}_e, 0)$ be an equilibrium point of (20) with $\|\mathbf{A}_e\| > 0$. Then, $\mathbf{A}_e$ must satisfy $f(0, C, \mathbf{A}_e) = 0$ or

$$4g(\|\mathbf{A}_e\|^2) = v^2(v^2 - 1) - 4C. \tag{34}$$

Suppose that $(\mathbf{A}_e, 0) \in \Gamma^1(0)$. Then, we must have

$$2U_1(\|\mathbf{A}_e\|^2, C) \triangleq \int_0^{\|\mathbf{A}_e\|^2} [\kappa g(\eta) + \gamma] d\eta = 0. \tag{35}$$

From (34), we have

$$2U_1(\|\mathbf{A}_e\|^2, C) = \int_0^{\|\mathbf{A}_e\|^2} \kappa[g(\eta) - g(\|\mathbf{A}_e\|^2)] d\eta, \tag{36}$$

which is a positive quantity for $\|\mathbf{A}_e\| > 0$ under condition (ii). This contradicts (35). Hence $(\mathbf{A}, \mathbf{A}') = (0,0)$ is the only equilibrium point in $\Gamma^1(0)$.

To show that $(\mathbf{A}, \mathbf{A}') = (0,0)$ is a saddle point, consider the following linearized equation (20) about $(\mathbf{A}, \mathbf{A}') = (0,0)$:

$$d^2\delta A_j/d\xi^2 = (\partial^2 U_1/\partial A_j^2)\big|_{\mathbf{A}=0} \delta A_j, \quad j = 1, \ldots, N, \tag{37}$$

where

$$(\partial^2 U_1/\partial A_j^2)\big|_{\mathbf{A}=0} = \gamma \tag{38}$$

Note that due to the symmetry of U_1 about $\mathbf{A} = 0$, $(\partial^2 U_1/\partial A_j \partial A_k)\big|_{\mathbf{A}=0} = 0$ for $j \neq k$. Under condition (i), $(\partial^2 U_1/\partial A_j^2)\big|_{\mathbf{A}=0} > 0$, $j = 1, \ldots, N$, so $(\mathbf{A}, \mathbf{A}') = (0,0)$ is a saddle point. Moreover, it is the limit point of some trajectory lying in $\Gamma^1(0)$ as $|\xi| \to \infty$. Hence, $C_1 = 0$ implies property (30).

Next, we must verify that there exist points $(\mathbf{A}(0), \mathbf{A}'(0)) = (\mathbf{A}(0), 0)$ with $\|\mathbf{A}(0)\| > 0$ such that $C_1 = 0$. From (15) and (31), this corresponds to finding a $\|\mathbf{A}(0)\| > 0$ such that

$$2U_1(\|\mathbf{A}(0)\|^2, C) \triangleq \int_0^{\|\mathbf{A}(0)\|^2} [\kappa g(\eta) + \gamma] d\eta = 0, \tag{39}$$

which, in view of (29), can be rewritten as

$$\|\mathbf{A}(0)\|^2 = W(\|\mathbf{A}(0)\|^2) \triangleq 4[v^2(v^2 - 1) - 4C] \times \int_0^{\|\mathbf{A}(0)\|^2} g(\eta)\, d\eta. \tag{40}$$

Evidently, under condition (ii), the mapping W has a nonzero fixed point $\|\mathbf{A}(0)\|^2 < \infty$.

We have established that there exist points $(\mathbf{A}(0), 0)$ in $\Gamma^1(0)$ with $\|\mathbf{A}(0)\| > 0$. Now, we must show that, for such a point, there exists a trajectory curve lying in $\Gamma^1(0)$ which joins $(\mathbf{A}(0), 0)$ and $(0,0)$. This is assured when $\Gamma^1(0)$ is compact. It is straightforward to show that $\Gamma^1(0)$ is closed. To show that $\Gamma^1(0)$ is bounded, we rewrite (32) as

$$w = \int_0^u [\kappa g(\eta) + \gamma] d\eta, \tag{41}$$

where $w = \|\mathbf{A}'\|^2$ and $u = \|\mathbf{A}\|^2$. Condition (i) implies that $\gamma > 0$ and $\kappa < 0$. From (ii), there exists a finite $u_1 > 0$ such that the right-hand side of (41) is zero at u_1 and negative for all $u > u_1$. Since (41) is valid only for $w \geq 0$, hence $\|\mathbf{A}\|^2 \leq u_1$. Also, from Weierstrass theorem, there exists a finite $W_1 > 0$ such that $\|\mathbf{A}'\|^2 \leq W_1$, since the right-hand side of (41) is continuous on the compact interval $0 \leq u \leq u_1$. Thus, the boundedness of $\Gamma^1(0)$ is established.

Finally, since only the nonnegative solutions of (20) are meaningful here, it remains to show that for a point $(\mathbf{A}(0), 0) \in \Gamma^1(0)$ with $A(0) \geq 0$ and $\|\mathbf{A}(0)\| > 0$, its corresponding solution is nonnegative, that is, $\mathbf{A}(\xi) \geq 0$ and for all $\xi \in \mathbb{R}$. This is immediately apparent from the fact that $\Gamma^1(0) = \Gamma_+^1(0) \cup \Gamma_-^1(0)$ and $\Gamma_+^1(0) \cap \Gamma_-^1(0) = \{(0,0)\}$, where $\Gamma_+^1(0) = \{(\mathbf{A}, \mathbf{A}') \in \Gamma^1(0) : A \geq 0$ and $\Gamma_-^1(0) = \{(\mathbf{A}, \mathbf{A}') \in \Gamma^1(0) : \mathrm{A} \leq 0\}$, since $U_1(0, C) = 0$. ∎

Remarks: (R-1) Theorems 1 and 2 give, respectively, sufficient conditions for the nonexistence and existence of multidimensional solitary-wave solutions of (1) which are directly analogous to those for the one-dimensional case. From (33), it is evident that $\Gamma^1(0)$ is symmetric about $\mathbf{A} = 0$ and $\mathbf{A}' = 0$. Also, $\Gamma_+^1(0)$ and $\Gamma_-^1(0)$ are symmetric about $\mathbf{A}' = 0$. Thus, under the conditions of Theorem 2, the trajectory curves in the $(\mathbf{A}, \mathbf{A}')$-space corresponding to the solitary-wave solutions of (7) satisfying (30) have similar properties, and they have the form:

$$h_j(\xi) = A_j(\xi) \exp\{i[\theta_j(0) + v\xi/2]\}, \quad j = 1, \ldots, N. \tag{42}$$

(R-2) Along any solution of (20), its corresponding density $s(\xi)$ can be found directly by solving (26) for $\|\mathbf{A}(\xi)\|^2$ or $|\mathbf{h}(\xi)|^2$ and substituting the result into (6) with $\hat{C} = 0$. Complete knowledge of the solution $\mathbf{A}(\xi)$ is not necessary here.

We note that if the assumptions of Theorem 2 are satisfied and there exists a positive number r_e such that

$$\kappa g(r_e^2) + \gamma = 0 \text{ or } g(r_e^2) = v^2(v^2 - 1)/4 - C, \tag{43}$$

then (20) has an uncountably infinite number of nonisolated, nonzero equilibrium points $(\mathbf{A}_e, 0)$ such that $\mathbf{A}_e$ lies on the sphere $\{\mathbf{A}_e \in \mathbb{R}^N : \|\mathbf{A}_e\| = r_e\}$.

Now, we show that there exist solutions $\mathbf{A}(\xi)$ of (20) in some neighborhood of these equilibrium points such that their norms are periodic functions of ξ.

First, we rewrite (22) in the form

$$d^2u/d\xi^2 = \partial V(u, C_1)/\partial u, \tag{44}$$

where

$$V(u, C_1) = 2\int_0^u \{\eta[\kappa g(\eta) + \gamma] + C_1 + 2U_1(\eta, C)\} d\eta, \tag{45}$$

and the initial conditions $u(0) = \|\mathbf{A}(0)\|^2$ and $u'(0)$

$= 2\mathbf{A}(0)\cdot\mathbf{A}'(0)$ are chosen such that condition (24) is satisfied. It can be readily verified that if we set $C_1 = C_1^0$ given by

$$C_1^0 = -\int_0^{r_e^2}[\kappa g(\eta)+\gamma]d\eta, \tag{46}$$

then $u_e = r_e^2$ is a stationary point of $V(\cdot, C_1^0)$, or $(u,u') = (r_e^2, 0)$ is an equilibrium point of (44). For $C_1 = C_1^0$, condition (24) becomes

$$\|\mathbf{A}'(0)\|^2 = \int_{r_e^2}^{\|\mathbf{A}(0)\|^2}[\kappa g(\eta)+\gamma]d\eta \geq 0. \tag{47}$$

Under condition (i) and (ii) of Theorem 2, we have $\kappa g(\eta)+\gamma > 0$ for $0 \leq \eta < r_e^2$, and $\kappa g(\eta)+\gamma < 0$ for all $\eta > r_e^2$. Evidently, (47) is satisfied if and only if $\|\mathbf{A}(0)\| = r_e$. Hence, the only solution to (44) with initial condition $(u(0), u'(0))$ satisfying (47) is the equilibrium solution $(u(\xi), u'(\xi)) = (r_e^2, 0)$ for all ξ.

Now, we consider the solutions of (44) with $C_1 = C_1^0 + \delta C_1$ and initial condition $(u(0), u'(0)) = (\|\mathbf{A}(0)\|^2, 2\mathbf{A}(0)\cdot\mathbf{A}'(0))$ satisfying (24), where δC_1 is a small perturbation of C_1 about C_1^0. Let $u_e(C_1)$ denote a stationary point of $V(\cdot, C_1)$ or a root of the equation

$$u[\kappa g(u)+\gamma] + C_1 + 2U_1(u, C) = 0. \tag{48}$$

For $C_1 = C_1^0 + \delta C_1$, we can write

$$u_e(C_1) = r_e^2 + \delta u_e. \tag{49}$$

Clearly, under the assumptions of Theorem 2, δu_e depends continuously on δC_1 and $|\delta u_e| \to 0$ as $|\delta C_1| \to 0$. Also, since

$$\partial^2 V(u, C_1)/\partial u^2 = 4[\kappa g(u)+\gamma] + 2\kappa u g'(u), \tag{50}$$

we have $(\partial^2 V(u, C_1)/\partial u^2)|_{u=r_e} = 2\kappa r_e g'(r_e) < 0$, or $u_e = r_e^2$ is a relative maximum point of $V(\cdot, C_1^0)$. In fact, since $2\kappa u g'(u) < 0$ for all $u > 0$ and $\kappa g(r_e^2)+\gamma = 0$, there exists a positive number ϵ such that for each δC_1, $|\delta C_1| < \epsilon$, it corresponding $u_e(C_1) = r_e^2 + \delta u_e$ is a relative maximum point of $V(\cdot, C_1^0 + \delta C_1)$. Consequently, for any fixed δC_1, $|\delta C_1| < \epsilon$, (44) has periodic solutions in some neighborhood of the corresponding equilibrium point $(u,u') = (r_e^2 + \delta u_e, 0)$.[11] They are given by the solution of

$$[u'(\xi)]^2/2 = V(u(\xi), C_1^0 + \delta C_1) - V(u_0, C_1^0 + \delta C_1) + (u_0')^2/2 \tag{51}$$

with $u(0) = u_0$, where the initial point $(u,(0), u'(0)) = (u_0, u_0')$ is sufficiently close to $(u, u') = (r_e^2 + \delta u_e, 0)$. In particular, we can choose $u_0 = \|\mathbf{A}(0)\|^2 > 0$ and $u_0' = 2\mathbf{A}(0)\cdot\mathbf{A}'(0)$ such that condition (24) is given by

$$\|\mathbf{A}'(0)\|^2 = C_1^0 + \delta C_1 + 2U_1(\|\mathbf{A}(0)\|^2, C) = \delta C_1 + \int_{r_e^2}^{\|\mathbf{A}(0)\|^2}[\kappa g(\eta)+\gamma]d\eta \geq 0 \tag{52}$$

is satisfied. This is possible for any positive δC_1. The existence of solutions of (20) in some neighborhood of the equilibrium points $(\mathbf{A}_e, 0)$ with $\|\mathbf{A}_e\| = r_e$, whose norms are periodic functions of ξ follows from the fact that $|\delta u_e| \to 0$ as $|\delta C_1| \to 0$. The foregoing result can be summarized as a theorem.

Theorem 3: Assume that the conditions of Theorem 2 are satisfied, and there exists a real number $r_e > 0$ satisfying (43). Then there exist solutions $\mathbf{A}(\xi)$ of (20) in some neighborhood of the equilibrium set $\{(\mathbf{A}, \mathbf{A}') \in \mathbb{R}^{2N}: \|\mathbf{A}\| = r_e, \mathbf{A}' = 0\}$ such that their norms $\|\mathbf{A}(\xi)\|$ are periodic functions of ξ.

Note that in the multidimensional case, the periodicity of $u(\xi) = \|\mathbf{A}(\xi)\|^2$ generally does not imply the periodicity of $\mathbf{A}(\xi)$. Since the energy density of the electric field is proportional to $|\mathbf{h}(\xi)|^2$, solutions with periodic $|\mathbf{h}(\xi)|$ represent oscillatory energy densities. Evidently, from (6) (with $\hat{C} = 0$), the periodicity of $s(\xi)$ is implied by that of $|\mathbf{h}(\xi)|$. Now, we give a simple sufficient condition for the nonexistence of periodic travelling waves in the sense that $|\mathbf{h}(\xi)|$ and $s(\xi)$ are periodic in ξ.

Theorem 4: Suppose that the follwing conditions are satisfied:

(i) g is a real-valued continuous monotone increasing function defined on $\mathbb{R}$ such that $g(0) = 0$;

(ii) $v^2(v^2-1) < 4C$ and $v^2 > 1$;

(iii) the initial conditions $\mathbf{A}(0)$ and $\mathbf{A}'(0)$ satisfy $\|\mathbf{A}(0)\| > 0$ and $\tilde{C}_1 \geq 0$, where $\tilde{C}_1$ is defined in (31).

Then, the norm of the corresponding solutions $\mathbf{A}(\xi)$ of (20) is nonperiodic in ξ.

Proof: Consider (22) given explicitly by

$$d^2u/d\xi^2 = 4\gamma u + 2\tilde{C}_1 + 2\kappa\{ug(u) + \int_0^u g(\eta)\,d\eta\}, \tag{53}$$

where γ and κ are as in (16). Under condition (i), the $\{\cdots\}$ term in (53) is nonnegative for $u \geq 0$. From conditions (ii) and (iii), we have $\gamma > 0$ so that $d^2u/d\xi^2 \geq 0$ for all $u \geq 0$. Since $u(0) = \|\mathbf{A}(0)\|^2 > 0$, u is a nonzero convex function of ξ which cannot be periodic. ∎

Theorem 4 gives a sufficient condition for the nonexistence of supersonic periodic travelling waves. In the subsonic case ($v^2 < 1$), the condition $v^2(v^2-1) < 4C$ implies that $\gamma < 0$. Thus, under condition (i) of Theorem 4, we have $d^2u/d\xi^2 \leq 0$ for all $u \geq 0$ when $\tilde{C}_1 \leq 0$, which implies the nonexistence of subsonic periodic travelling waves. However, $\tilde{C}_1 \leq 0$ corresponds to

$$\|\mathbf{A}'(0)\|^2 \leq \gamma\|\mathbf{A}(0)\|^2 + \kappa\int_0^{\|\mathbf{A}(0)\|^2} g(\eta)\,d\eta, \tag{54}$$

whose right-hand side is nonpositive. Thus, this condition can be satisfied only in the trivial case when $\mathbf{A}(0) = 0$ and $\mathbf{A}'(0) = 0$.

3. SPECIAL CASES

Now, we apply the results in Sec. 2 to Eq. (1) with particular forms of g arising in physical situations.

3.1: $g(|\mathbf{E}|^2) = |\mathbf{E}|^2$: This corresponds to the case with the usual pondermotive force. Here, U_1 as given by (15) has the explicit form:

$$2U_1(\|\mathbf{A}\|^2, C) = \gamma\|\mathbf{A}\|^2 + \kappa\|\mathbf{A}\|^4/2, \tag{55}$$

where γ and κ are as in (16). A first integral of (13) is given by

$$\|\mathbf{A}'(\xi)\|^2 - \gamma\|\mathbf{A}(\xi)\|^2 - \kappa\|\mathbf{A}(\xi)\|^4/2 - \sum_{j=1}^{N}\mu_j^2 A_j^{-2}(\xi)/4 = C_1. \tag{56}$$

When $\mathbf{A}(0)$ and $\theta(0)$ are chosen such that $\mu = 0$, the

equation for $A_j(\xi)$ given by (20) reduces to

$$d^2A_j/d\xi^2 = (\gamma + \kappa\|\mathbf{A}\|^2)A_j, \quad j = 1, \ldots, N, \tag{57}$$

and the equation for $u(\xi) = \|\mathbf{A}(\xi)\|^2$ given by (22) becomes

$$d^2u/d\xi^2 = 3\kappa u^2 + 4\gamma u + 2C_1. \tag{58}$$

A first integral of (53) is given by

$$(u'(\xi))^2/2 = \kappa u^3(\xi) + 2\gamma u^2(\xi) + 2C_1 u(\xi) + C_2, \tag{59}$$

where C_2 is an integration constant. By restricting the right-hand side of (59) to be nonnegative, we have the following implicit expression for $u(\xi)$:

$$\int_{u(0)}^{u(\xi)} (\kappa\eta^3 + 2\gamma\eta^2 + 2C_1\eta + C_2)^{-1/2} d\eta = \sqrt{2}\xi, \quad \xi \in \mathbb{R}. \tag{60}$$

Now, we apply Theorem 1 to this special case. Clearly, for $v^2 < 1$, condition (27) cannot be satisfied. But, for $v^2 > 1$, (27) is satisfied when $v^2(v^2-1)/4 = C$. Under this condition, there do not exist solutions of (20) or solitary-wave solutions of (7) such that $|\mathbf{h}(0)| > 0$ and $|\mathbf{h}(\xi)| \to 0$ as $|\xi| \to \infty$. To apply Theorem 2 to this special case, we see that under condition (i), $g(u) = u$ satisfies condition (ii). Thus, under condition (i), (57) has solitary-wave solutions satisfying (30). For such solutions, we see from (56) with $\mu = 0$ that C_1 must be zero. Also, from (59), C_2 must be zero. Since $u(0) = \|\mathbf{A}(0)\|^2 > 0$ and $u'(0) = 0$, $\|\mathbf{A}(0)\|$ must satisfy

$$\|\mathbf{A}(0)\|^2 = -2\gamma/\kappa, \tag{61}$$

where $\gamma > 0$ and $\kappa < 0$ under condition (i). Thus, (60) becomes

$$\int_{2\gamma/|\kappa|}^{\|\mathbf{A}(\xi)\|^2} (\kappa\eta^3 + 2\gamma\eta^2)^{-1/2} d\eta = \sqrt{2}\,\xi, \quad \xi \in \mathbb{R}. \tag{62}$$

It follows that

$$\|\mathbf{A}(\xi)\| = (2\gamma/|\kappa|)^{1/2} \operatorname{sech}(\sqrt{\gamma}\xi), \quad \xi \in \mathbb{R}, \tag{63}$$

and, in view of (6), we have

$$s(\xi) = (v^2-1)^{-1}[(2\gamma/|\kappa|)\operatorname{sech}^2(\sqrt{\gamma}\,\xi) + C]. \tag{64}$$

Substituting (64) into (11) with $\mu_j = 0$ leads to a set of uncoupled equations for A_j's given given by

$$d^2A_j/d\xi^2 = (\gamma + \kappa\|\mathbf{A}\|^2)A_j = \gamma[1 - 2\operatorname{sech}^2(\sqrt{\gamma}\,\xi)]A_j, \quad j = 1, \ldots, N, \tag{65}$$

which can be integrated independently to obtain $A_j(\xi)$. The foregoing results are consistent with those for the one-dimensional solitary waves.[10]

Turning now to the periodic travelling waves, we observe that, under condition (i) of Theorem 2, (57) has an uncountably infinite number of nonisolated equilibrium points $(\mathbf{A}_e, 0)$ such that $\|\mathbf{A}_e\|^2 = r_e^2 = -\gamma/\kappa > 0$. Also, (58) can be written in the form of (44) with

$$V(u, C_1) = u(\kappa u^2 + 2\gamma u + 2C_1). \tag{66}$$

If we set $C_1 = C_1^0 \triangleq \gamma^2/(2\kappa)$, then $u = -\gamma/\kappa$ is a relative maximum point of $V(\cdot, C_1^0)$. Now, we consider the solutions of (58) with initial conditions $u(0) = \|\mathbf{A}(0)\|^2$ and $u'(0) = 2\mathbf{A}(0)\cdot\mathbf{A}'(0)$ satisfying condition (24) given explicitly by

$$\|\mathbf{A}'(0)\|^2 = C_1 + \gamma u(0) + \kappa u^2(0)/2 \geq 0. \tag{67}$$

Let $\mathcal{U}(C_1)$ denote the set of all $u(0) \geq 0$ satisfying (67) for a fixed C_1. It can be readily verified that, under condition (i) of Theorem 2, we have

$$\mathcal{U}(C_1) = \{u(0): [\gamma - (\gamma^2 + 2|\kappa|C_1)^{1/2}]/|\kappa| \leq u(0) \leq [\gamma + (\gamma^2 + 2|\kappa|C_1)^{1/2}]/|\kappa|\} \text{ for } 0 \geq C_1 \geq \gamma^2/(2\kappa) \tag{68}$$

$$\mathcal{U}(C_1) = \{u(0): 0 \leq u(0) \leq [\gamma + (\gamma^2 + 2|\kappa|C_1)^{1/2}]/|\kappa|\} \text{ for } C_1 \geq 0, \tag{69}$$

and $\mathcal{U}(C_1)$ is empty for $C_1 < \gamma^2/(2\kappa)$. Note that for $C_1 = \gamma^2/(2\kappa)$, $\mathcal{U}(C_1)$ contains only the point $u(0) = -\gamma/\kappa$. Thus, from Theorem 3, if condition (i) of Theorem 2 is satisfied, then there exist solutions $\mathbf{A}(\xi)$ of (20) in some neighborhood of the equilibrium set $\{(\mathbf{A}, \mathbf{A}') \subset \mathbb{R}^{2N}: \|\mathbf{A}\|^2 = -\gamma/\kappa, \mathbf{A}' = 0\}$ such that their norms $\|\mathbf{A}(\xi)\|$ are periodic in ξ. These solution curves correspond to (56) with $\mu = 0$ and C_1 satisfying $0 > C_1 > \gamma^2/(2\kappa)$. When C_1 is set to zero, we have solitary-wave solutions such that $\|\mathbf{A}(\xi)\|$ and $\|\mathbf{A}'(\xi)\| \to 0$ as $|\xi| \to \infty$ as given by (63). In this case, $(u, u') = (0, 0)$ is a saddle point of (58) with $C_1 = 0$.

Figure 1 shows the trajectories of (58) with $\gamma = 1$ and $\kappa = -2$ in the (u, u')-plane for various values of C_1 and $u(0)$ satisfying (67). Note that, for $C_1 > 0$, $(u, u') = (0, 0)$ is not an equilibrium point of (58). In fact, these solutions pass through the origin and they are periodic functions of ξ. Figure 2 shows the behavior of the trajectories in the ξ domain.

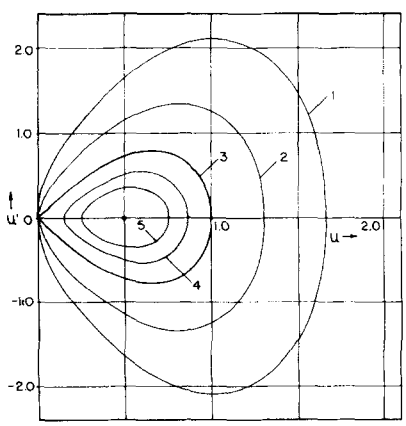


FIG. 1. Trajectories of (58) with $\gamma = 1$ and $\kappa = -2$ in the (u, u') plane for $u(0)$ satisfying (76) and fixed values of C_1, (curves 1–5 correspond to $C_1 = 1.1,\ 0.4,\ 0.0,\ -1/8,\ -3/16$, respectively); curve 3 is the solitary-wave solution.

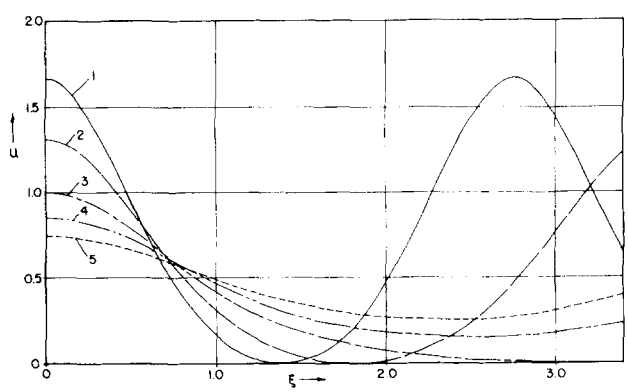


FIG. 2. Behavior of the trajectories shown in Fig. 1 in the ξ domain.

Finally, for the supersonic case $v^2 > 1$, we have from Theorem 4 that if $v^2(v^2-1) < 4C$ and $\mathbf{A}(0)$, $\mathbf{A}'(0)$ satisfy $\|\mathbf{A}(0)\| > 0$ and $\tilde{C}_1 \geq 0$ or

$$\|\mathbf{A}'(0)\|^2 \geq \gamma\|\mathbf{A}(0)\|^2 + \kappa\|\mathbf{A}(0)\|^4/2 > 0, \tag{70}$$

then the norm of the corresponding solution $\mathbf{A}(\xi)$ of (57) is nonperiodic in ξ.

3.2: $g(|\mathbf{E}|^2) = K[1 - \exp(-|\mathbf{E}|^2)]$: This form of g, with K being a positive constant, has been proposed by Wilcox and Wilcox[12] to represent ion density saturation. For this g, U_1 is given by

$$2U_1(\|\mathbf{A}\|^2, C) = (\kappa K + \gamma)\|\mathbf{A}\|^2 - \kappa K[1 - \exp(-\|\mathbf{A}\|^2)], \tag{71}$$

and a first integral of (13) is given by

$$\|\mathbf{A}'(\xi)\|^2 - (\kappa K + \gamma)\|\mathbf{A}(\xi)\|^2 + \kappa K[1 - \exp(-\|\mathbf{A}(\xi)\|^2)] - \sum_{j=1}^{N} \mu_j^2 A_j^2(\xi)/4 = C_1. \tag{72}$$

The equations for the $A_j(\xi)$'s with $\mu = 0$ corresponding to (20) have the form

$$d^2A_j/d\xi^2 = (\gamma + \kappa K\{1 - \exp[-\|\mathbf{A}(\xi)\|^2]\})A_j, \quad j = 1, \ldots, N. \tag{73}$$

The evolution of $u(\xi) = \|\mathbf{A}(\xi)\|^2$ with ξ along an integral curve of (73) specified by C_1 is governed by

$$d^2u/d\xi^2 = 2\{2(\gamma + \kappa K)u + C_1 - \kappa K[1 - (1-u)\exp(-u)]\}, \tag{74}$$

which has a first integral of the form

$$[u'(\xi)]^2/2 = 2(\gamma + \kappa K)u^2(\xi) + 2u(\xi)(C_1 + \kappa K\{\exp[-u(\xi)] - 1\}) + C_2, \tag{75}$$

where C_2 is an integration constant. By restricting the right-hand side of (75) to be nonnegative, we can integrate (75) to give an implicit expression for $u(\xi)$.

To apply Theorem 1 to this case, consider inequality (27) given explicitly by

$$\kappa K[1 - \exp(-u)] + \gamma > 0 \quad \text{for all } u \geq 0. \tag{76}$$

This condition is satisfied when

$$v^2 > 1 \quad \text{and } \gamma > 0 \tag{77}$$

or

$$v^2 < 1 \quad \text{and } \kappa K + \gamma > 0. \tag{78}$$

Thus, under (77) or (78), there do not exist solitary-wave solutions such that $|\mathbf{h}(0)| > 0$ and $|\mathbf{h}(\xi)| \to 0$ as $|\xi| \to \infty$. Moreover, from Theorem 4, when $v^2 > 1$, $\gamma > 0$ and $(\mathbf{A}(0), \mathbf{A}'(0))$ satisfies

$$\|\mathbf{A}'(0)\|^2 \geq (\kappa K + \gamma)\|\mathbf{A}(0)\|^2 - \kappa K\{1 - \exp[-\|\mathbf{A}(0)\|^2]\}, \tag{79}$$

then the norm of the corresponding solution $\mathbf{A}(\xi)$ of (73) is nonperiodic in ξ.

Now, consider condition (29) in Theorem 2 which requires the existence of a $u_1 > 0$ such that

$$K[u_1 + \exp(-u_1) - 1] = (-\gamma/\kappa)u_1 \tag{80}$$

and, for all $u > u_1$,

$$K[u + \exp(-u) - 1] > (-\gamma/\kappa)u. \tag{81}$$

This condition is satisfied if

$$\gamma/\kappa < 0. \tag{82}$$

Thus, if $v^2 < 1$ and $\gamma > 0$, then the hypotheses of Theorem 2 are satisfied. Hence (73) has solitary-wave solutions satisfying (30). For such a solution with $\|\mathbf{A}'(0)\| = 0$, we have from (72) with $C_1 = 0$ and $\mu = 0$ that $\|\mathbf{A}(0)\|$ must satisfy

$$\|\mathbf{A}(0)\|^2 = \kappa K(\kappa K + \gamma)^{-1}\{1 - \exp[-\|\mathbf{A}(0)\|^2]\}, \tag{83}$$

which always has a solution $\|\mathbf{A}(0)\|^2 > 0$ if $v^2 < 1$ and $\gamma > 0$.

Next, we observe that if

$$\kappa K/(\gamma + \kappa K) > 1, \tag{84}$$

then (73) has an uncountably infinite number of nonisolated equilibrium points $(\mathbf{A}, \mathbf{A}') = (\mathbf{A}_e, 0)$ such that

$$u_e \triangleq \|\mathbf{A}_e\|^2 = \ln[\kappa K/(\gamma + \kappa K)] > 0. \tag{85}$$

Note that if $v^2 < 1$, then (84) implies (82) and

$$(\gamma + \kappa K) < 0. \tag{86}$$

We shall verify that, under the conditions of Theorem 2, (74) has periodic solutions in some neighborhood of the point $(u, u') = (u_e, 0)$.

First, we rewrite (74) in the form of (44) with V given by

$$V(C_1, u) = 2(\gamma + \kappa K)u^2 + 2(C_1 - \kappa K)u + 2\kappa K u \exp(-u). \tag{87}$$

If we set $C_1 = C_1^0$ given by

$$C_1^0 = 2(\gamma + \kappa K)\ln\left[\frac{(\gamma + \kappa K)}{(\kappa K)}\right] + \kappa K - (\gamma + \kappa K)\left[1 - \ln\left(\frac{\kappa K}{(\gamma + \kappa K)}\right)\right], \tag{88}$$

then u_e given by (85) is a stationary point of $V(\cdot, C_1^0)$, or $(u, u') = (u_e, 0)$ is an equilibrium point of (74). At this point, $[\partial^2 V(u, C_1^0)/\partial u^2]|_{u=u_e} = 2u_e(\gamma + \kappa K)$. Thus, under condition (86), u_e is a relative maximum point of $V(\cdot, C_1^0)$. Now we consider the solutions of (74) for various values of C_1 in some neighborhood of C_1^0, with initial conditions $u(0) = \|\mathbf{A}(0)\|^2$ and $u'(0) = 2\mathbf{A}(0) \cdot \mathbf{A}'(0)$ satisfying condition (24) given by

$$\|\mathbf{A}'(0)\|^2 = C_1 + (\gamma + \kappa K)u(0) - \kappa K\{1 - \exp[-u(0)]\} \geq 0. \tag{89}$$

As in Sec. 3.1, let $\mathcal{U}(C_1)$ denote the set of all $u(0) \geq 0$ satisfying (89) for a fixed C_1 or

$$\mathcal{U}(C_1) = \{u(0) \geq 0: C_1 + (\gamma + \kappa K)u(0) \geq \kappa K[1 - \exp[-u(0)]]\}. \tag{90}$$

It can be readily verified that if $v^2 < 1$ and condition (84) is satisfied, then $\mathcal{U}(C_1)$ is empty for all $C_1 < C_1^*$, and

$$\mathcal{U}(C_1^*) = \{u^*(0)\}, \tag{91}$$

where

$$u^*(0) = \ln[\kappa K/(\gamma + \kappa K)], \tag{92}$$

$$C_1^* = (\gamma + \kappa K)\ln[\kappa K/(\gamma + \kappa K)], \tag{93}$$

where $u^*(0)$ corresponds to the point of tangency between the line $y_1(u) = C_1^*(\kappa K)^{-1} + [1 + \gamma(\kappa K)^{-1}]u$ and the curve $y_2(u) = 1 - \exp(-u)$. Also,

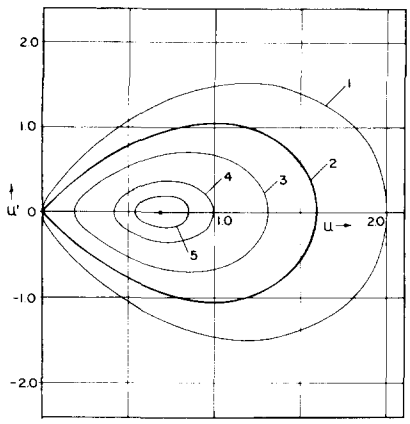


FIG. 3. Trajectories of (74) with $\gamma=1$ and $\kappa=-2$ in the (u,u') plane for $u(0)$ satisfying (89) and fixed values of C_1 (curves 1—5 correspond to $C_1=0.2707$, 0.0, -0.1522, -0.2642, -0.3069 respectively); curve 2 is the solitary-wave solution.

$$\mathcal{U}(C_1)=\{u(0):\ \check{u}\le u(0)\le\hat{u}\}\ \text{ for } C_1^*\le C_1\le 0, \tag{94}$$

$$\mathcal{U}(C_1)=\{u(0):\ 0\le u(0)\le\hat{u}\}\ \text{ for } C_1>0, \tag{95}$$

where $\check{u}$ and $\hat{u}$ with $\check{u}<\hat{u}$ are the two distinct positive roots of the equation

$$C_1+(\gamma+\kappa K)u-\kappa K[1-\exp(-u)]=0. \tag{96}$$

Thus, from Theorem 3, if $v^2<1$ and (84) is satisfied, then there exist solutions $\mathbf{A}(\xi)$ of (73) in some neighborhood of the equilibrium set $\{(\mathbf{A},\mathbf{A}')\in\mathbb{R}^{2n}:\ \|\mathbf{A}\|^2=\ln[\kappa K/(\gamma+\kappa K)],\ \mathbf{A}'=\mathbf{0}\}$ such that their norms $\|\mathbf{A}(\xi)\|$ are periodic in ξ. These solution curves correspond to (72) with $\mu=0$ and C_1 satisfying $0>C_1>C_1^*$. The trajectories of (74) with $\gamma=1$, $\kappa=-2$, and $K=1$ for various values of C_1 and $u(0)$ satisfying (89) are shown in Fig. 3. Their corresponding trajectories in the ξ domain are shown in Fig. 4.

4. CONCLUSIONS

We have shown that, under mild conditions on the nonlinearity g, (1) has multidimensional solitary-wave and periodic travelling-wave solutions $(\mathbf{E}(\xi),n(\xi))$ in the sense that $|\mathbf{E}(\xi)|$ and $n(\xi)$ tend to finite values as $|\xi|\to\infty$, and they are periodic functions of ξ respectively. Along these solutions, the phase of $\mathbf{E}(\xi)$ is an affine function of ξ. Moreover, $u(\xi)=|\mathbf{E}(\xi)|^2$ satisfies a scalar second-order ordinary differential equation whose solutions have properties similar to those in the one-dimensional case. Although in this study, we have treated only the case with electrostatic

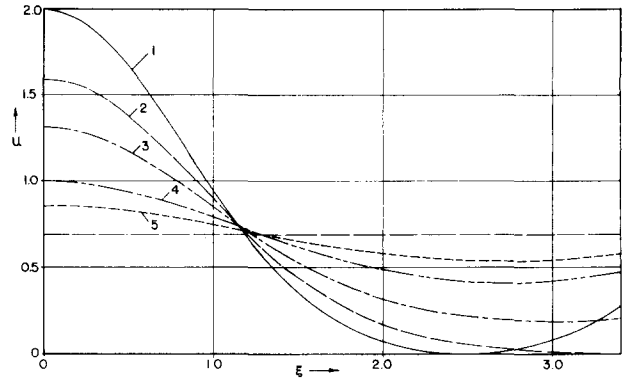


FIG. 4. Behavior of the trajectories shown in Fig. 3 in the ξ domain.

waves (i.e., $\nabla\times\mathbf{E}=0$), the same approach may be used to obtain results for electromagnetic waves.

ACKNOWLEDGMENT

This work was supported by an AFOSR Grant No. 74-2662.

[1]V.E. Zakharov, Zh. Eksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].
[2]V.D. Shapiro, Zh. Eksp. Teor. Fiz. **66**, 945 (1974) [Sov. Phys. JETP **39**, 459 (1974)].
[3]L.M. Degtyarev, V.G. Nakhan'kov, and L.I. Rudakov, Zh. Eksp. Teor. Fiz. **67**, 533 (1974) [Sov. Phys. JETP **40**, 264 (1975)].
[4]L.M. Degtyarev, V.E. Zakharov, and L.I. Rudakov, Zh. Eksp. Teor. Fiz. **68**, 115 (1975) [Sov. Phys. JETP **41**, 57 (1975)].
[5]V.E. Zakharov, A.F. Mastryukov, and V.S. Synakh, Sov. J. Plasma Phys. 1, 339 (1975).
[6]Kh. O. Abdulloev, I.L. Bogolyubskij, and V.G. Makhan'kov, Nucl. Fusion **15**, 21 (1975).
[7]V.V. Gorev, A.S. Kingsep, and V.V. Yan'kov, Zh. Eksp. Teor. Fiz. **70**, 921 (1976) [Sov. Phys. JETP **43**, 479 (1976)].
[8]V.N. Tsytovich, in *Plasma Physics, Nonlinear Theory and Experiments*, edited by H. Wilhelmsson (Plenum, New York, 1977), p. 166.
[9]K. Kato, K. Mima, M. Watanabe, and K. Nishikawa, in *Plasma Physics, Nonlinear Theory and Experiments*, edited by H. Wilhelmsson (Plenum, New York, 1977), p. 142.
[10]J. Gibbons *et al.*, J. Plasma Phys. **17**, pt. 2, 153 (1977).
[11]J. Hale, *Ordinary Differential Equations* (Wiley-Interscience, New York, 1969).
[12]J.Z. Wilcox and T.J. Wilcox, Phys. Rev. Lett. **37**, 1160 (1975).

Optical analysis of resonances in a velocity-dependent potential [a)]

J. Sesma

Departamento de Física Teórica, Facultad de Ciencias, and Instituto de Física Nuclear y de Altas Energías, Universidad de Zaragoza, Zaragoza, Spain

V. Vento

Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valencia, Valencia, Spain
(Received 3 November 1977)

Resonances in a velocity-dependent square well potential are analyzed in terms of multiple reflections of the incident wave on the internal walls of the potential well. The results differ considerably from those obtained for static potentials with a similar analysis.

1. INTRODUCTION

Resonance phenomena are familiar in several branches of physics. Many authors have studied their peculiarities. McVoy, Heller, and Bolsterli[1] have analyzed a variety of S-wave scattering problems. Others have considered resonances in potential scattering[2] and related topics such as residues at resonances,[3] time delay,[4] and resonance wavefunctions.[5]

In a series of previous papers[6,7] we have studied the analyticity of the scattering matrix for a velocity-dependent potential. Our study has put in evidence the existence, for such potential, of very characteristic resonances. The purposes of this paper is to discuss these resonances in the velocity-dependent potential from an optical point of view (i.e., in terms of multiple reflections), in order to obtain a better understanding of the physics involved in these phenomena. The method used is close to that developed by McVoy *et al.*[1] for a static potential, although our analysis runs along a different line and is not limited to S wave.

Let us consider the scattering of a particle of mass m and energy E by the velocity-dependent potential

$$V(\mathbf{r},\mathbf{p}) = A\mathbf{p}\cdot\Theta(b-r)\mathbf{p}/2m \tag{1.1}$$

introduced by Razavy, Field, and Levinger[8] in the description of nuclear forces. Let us call k and k' the wavenumbers respectively outside and inside the potential well,

$$k = (2mE)^{1/2}/\hbar, \quad k' = k(1+A)^{-1/2}. \tag{1.2}$$

In our subsequent analysis we shall use the adimensional wavenumbers

$$\alpha = kb, \quad \beta = k'b, \tag{1.3}$$

where b represents the range of the square potential. For our potential, partial wave analysis for the scattering matrix[6,7] leads us to

$$S_l(k) = -\frac{\alpha h_l^{(2)\prime}(\alpha) j_l(\beta) - (1+A)\beta h_l^{(2)}(\alpha) j_l'(\beta)}{\alpha h_l^{(1)\prime}(\alpha) j_l(\beta) - (1+A)\beta h_l^{(1)}(\alpha) j_l'(\beta)}. \tag{1.4}$$

Here, j_l, $h_l^{(1)}$, and $h_l^{(2)}$ are the familiar spherical Bessel and Hankel functions and the primes stand for derivatives with respect to their arguments.

The analysis of the Regge trajectories[7] for the potential given by Eq. (1.1) has allowed us to recognize the existence of resonances, at certain energies, for different ranges of values of the intensity parameter A. In the case of a velocity-dependent barrier ($0 < A$), the modulus of the scattering amplitude shows in each partial wave an infinity of bumps. These bumps are, nevertheless, very broad and cannot be considered as resonances. For a "shallow" velocity-dependent well ($-1 < A < 0$), an infinite number of resonances occur for each partial wave. The resonances become sharper and sharper as A tends to -1. If the potential well is "moderately deep" ($-2 < A < -1$), only one resonance appears for each partial wave. Such resonances become narrower as A approaches -2. Finally, for a "very deep" well ($A < -2$) there is one resonance only for values of the angular momentum l such that $l < (-2 - A)^{-1}$. For the highest physical value of l fulfilling this restriction, the resonance is very sharp. For smaller values of l, it becomes broader and broader. Moreover, these resonances for $A < -2$ are associated with a Regge pole which for zero energy is located at $l = (-2 - A)^{-1}$ in the angular momentum plane and which leaves the real l-axis moving to the left and downwards as the energy increases, following a Regge trajectory in the fourth quadrant of the l plane.

In Sec. 2 we define, following Nussenzveig,[9] coefficients for reflection on and transmission across the spherical surface of radius b separating the interior and the exterior of the potential. We rewrite the S matrix in terms of these coefficients and obtain in this way its Debye expansion. Sections 3 and 4 are devoted to a physical explanation of the resonances found for the potential Eq. (1.1), in the cases $-1 < A$ and $A < -1$ respectively, in terms of the reflection and transmission coefficients defined in Sec. 2.

2. THE DEBYE EXPANSION

The scattering by a cutoff potential can be explained in terms of interactions with the discontinuity surface of the potential well or barrier. What one obtains in this way is a parallel with geometrical optics.[1,2,9] In our subsequent treatment of the scattering by the velo-

[a)] This work has been supported by Instituto de Estudios Nucleares.

 0022-2488/78/1906-1293$1.00

city-dependent potential of Eq. (1.1), we follow closely, with the necessary modifications, the procedure developed by Nussenzveig[9] in the analysis of high frequency scattering by a transparent sphere.

Our potential is different from zero only in the interior of a sphere of radius b. Let us design by 1 this region. In the outside region of the sphere, which we denote by 2, there is no interaction. Since the potential possesses spherical symmetry, we may study separately each angular momentum partial wave.

Let us consider an incoming spherical wave of angular momentum l which arrives at the surface of radius b coming from region 2. This wave will be partly reflected at the potential discontinuity and partly transmitted to region 1. The radial wavefunction can be written

$$\Psi_{2,l}(r)=h_l^{(2)}(kr)/h_l^{(2)}(\alpha)+R_{22}h_l^{(1)}(kr)/h_l^{(1)}(\alpha), \tag{2.1}$$

$$\Psi_{1,l}(r)=T_{21}h_l^{(2)}(k'r)/h_l^{(2)}(\beta). \tag{2.2}$$

Here, R_{22} and T_{21} are, respectively, the reflection and transmission coefficients for the wave incoming from region 2. Of course, both coefficients depend on the energy and angular momentum of the particle. The explicit form of such dependence can be obtained from the continuity conditions satisfied by the wavefunction and its radial derivative. For our velocity-dependent potential these conditions are[6]

$$\Psi_{2,l}(b)=\Psi_{1,l}(b), \tag{2.3}$$

$$\Psi'_{2,l}(b)=(1+A)\Psi'_{1,l}(b). \tag{2.4}$$

Let us denote

$$\lambda=l+1/2, \tag{2.5}$$

as usual. Following Nussenzveig,[9] we introduce, for the sake of brevity, the symbols

$$[z]=J'_\lambda(z)/J_\lambda(z), \tag{2.6}$$

$$[1z]=H_\lambda^{(1)\prime}(z)/H_\lambda^{(1)}(z), \tag{2.7}$$

$$[2z]=H_\lambda^{(2)\prime}(z)/H_\lambda^{(2)}(z), \tag{2.8}$$

to design the logarithmic derivatives of the cylindrical Bessel and Hankel functions. With this notation, we obtain from Eqs. (2.3) and (2.4)

$$R_{22}(\lambda,E)=-\frac{\alpha[2\alpha]-(1+A)\beta[2\beta]+A/2}{\alpha[1\alpha]-(1+A)\beta[2\beta]+A/2}, \tag{2.9}$$

$$T_{21}(\lambda,E)=\frac{\alpha[1\alpha]-\alpha[2\alpha]}{\alpha[1\alpha]-(1+A)\beta[2\beta]+A/2}. \tag{2.10}$$

In a similar way, by considering an outgoing spherical wave of angular momentum l which arrives, from region 1, at the surface of radius b, we obtain for the reflection and transmission coefficients

$$R_{11}(\lambda,E)=-\frac{\alpha[1\alpha]-(1+A)\beta[1\beta]+A/2}{\alpha[1\alpha]-(1+A)\beta[2\beta]+A/2}, \tag{2.11}$$

$$T_{12}(\lambda,E)=(1+A)\frac{\beta[1\beta]-\beta[2\beta]}{\alpha[1\alpha]-(1+A)\beta[2\beta]+A/2}. \tag{2.12}$$

It is interesting to notice, from Eqs. (2.9)–(2.12), that there is no difficulty in extending the definition of reflection and transmission coefficients to unphysical (i.e., complex) values of the angular momentum and the energy. Then it is easy to check, from the definitions Eqs. (2.9)–(2.12) and very general properties of the Hankel function (see Ref. 10, Eq. 9.1.6), that all the coefficients are even functions of λ:

$$R_{ii}(-\lambda,E)=R_{ii}(\lambda,E), \quad i=1,2, \tag{2.13}$$

$$T_{ij}(-\lambda,E)=T_{ij}(\lambda,E), \quad i,j=1,2. \tag{2.14}$$

Useful relations involving these coefficients can be obtained from Eqs. (2.3), (2.4) and their complex conjugates, or directly from the expressions Eqs. (2.9)–(2.12). For instance,

$$T_{12}=1+R_{11}, \tag{2.15}$$

$$T_{21}=1+R_{22}, \tag{2.16}$$

$$\mathrm{Im}\{(1+A)^{-1}|T_{12}|^2\alpha[1\alpha]-\beta[1\beta]-|R_{11}|^2\beta[2\beta]$$

$$-R_{11}(\beta[2\beta]-\overline{\beta[1\beta]})\}=0, \tag{2.17}$$

$$\mathrm{Im}\{(1+A)|T_{21}|^2\beta[2\beta]-\alpha[2\alpha]-|R_{22}|^2\alpha[1\alpha]$$

$$-R_{22}(\alpha[1\alpha]-\overline{\alpha[2\alpha]})\}=0, \tag{2.18}$$

where the bar denotes complex conjugate. In the case of λ and E being real, the last two equations admit a more concise form in view of the relation

$$\overline{z[2z]}=z[1z], \quad \lambda, z, \text{ real}. \tag{2.19}$$

We shall return later on this point.

It is not difficult to write the S matrix, given by Eq. (1.4), in terms of the reflection and transmission coefficients. It becomes

$$S(\lambda,E)=[H_\lambda^{(2)}(\alpha)/H_\lambda^{(1)}(\alpha)]$$
$$\times$$
$$\{R_{22}+T_{21}[H_\lambda^{(1)}(\beta)/H_\lambda^{(2)}(\beta)](1-\rho)^{-1}T_{12}\}, \tag{2.20}$$

where we have denoted

$$\rho(\lambda,E)=R_{11}H_\lambda^{(1)}(\beta)/H_\lambda^{(2)}(\beta). \tag{2.21}$$

It is to be noticed that the above expression for the S matrix is exactly the same as for a static (i.e., velocity-nondependent) potential,[9] although the expressions for the reflection and transmission coefficients are unavoidably different in the velocity-dependent and static cases. The Debye expansion turns out by writing as a geometric series the factor $(1-\rho)^{-1}$ on the right-hand side of Eq. (2.20). We obtain

$$S(\lambda,E)=[H_\lambda^{(2)}(\alpha)/H_\lambda^{(1)}(\alpha)]$$
$$\times\left(R_{22}+T_{21}[H_\lambda^{(1)}(\beta)/H_\lambda^{(2)}(\beta)]\sum_{p=0}^{\infty}\rho^pT_{12}\right), \tag{2.22}$$

as in the static case. This expression provides a very simple explanation[9] of the scattering process: For an incident spherical wave, an interaction takes place at $r=b$ (factor $H_\lambda^{(2)}(\alpha)/H_\lambda^{(1)}(\alpha)$) and the wave is directly reflected (term R_{22}) or transmitted into the sphere (factor T_{21}); the transmitted part goes back and forth in the interior of the sphere (factor $[H_\lambda^{(1)}(\beta)/H_\lambda^{(2)}(\beta)]^{p+1}$) being totally reflected $p+1$ times at $r=0$ and partially

reflected p times (factor R_{11}^p) at the interior of the surface, to be finally transmitted to the outside (factor T_{12}). Of course, the validity of the Debye expansion is restricted by the convergence of the geometrical series, that is, it requires the condition $|\rho|<1$ to be satisfied. This condition will be examined in Secs. 3 and 4.

There are several attributes which allow to recognize a scattering resonance. One of them is the maximum, at the resonant energy, shown by the modulus of the amplitude of the internal or trapped wave. In view of the analogy with the optical problem provided by the Debye expansion, it is clear that the amplitude of the internal wavefunction is given by

$$\mathcal{A}(\lambda,E)=T_{21}[H_\lambda^{(1)}(\beta)/H_\lambda^{(2)}(\beta)](1-\rho)^{-1}. \tag{2.23}$$

In what follows we discuss the possible existence of maxima of $\mathcal{A}$ for positive energies and physical values of the angular momentum.

3. CASE OF A BARRIER OR A "SHALLOW" WELL

Let us start with the case of $0<A$ (barrier) or $-1<A<0$ ("shallow" well) in Eq. (1.1). At positive energies both α and β are real and Eq. (2.19) can be applied with z replaced by α and β. Bearing this in mind, it is easy to obtain

$$|R_{11}|^2+(1+A)^{-1}|T_{12}H_\lambda^{(1)}(\beta)/H_\lambda^{(1)}(\alpha)|^2=1, \tag{3.1}$$

$$|R_{22}|^2+(1+A)|T_{21}H_\lambda^{(2)}(\alpha)/H_\lambda^{(2)}(\beta)|^2=1, \tag{3.2}$$

from Eqs. (2.17) and (2.18) and the value of the Wronskian of $H_\lambda^{(1)}$ and $H_\lambda^{(2)}$ (see Ref. 10, Eq. 9.1.17). Equations (3.1) and (3.2) express the conservation of the probability in the interaction of the outgoing and incoming waves with the surfaces of the potential. They can be written also in the form

$$|R_{11}|^2+\overline{T_{12}}T_{21}=1, \tag{3.3}$$

$$|R_{22}|^2+T_{12}\overline{T_{21}}=1. \tag{3.4}$$

It is evident that $\overline{T_{12}}T_{21}$ is a real quantity, that is,

$$\overline{T_{12}}T_{21}=T_{12}\overline{T_{21}}. \tag{3.5}$$

This implies

$$|R_{11}|=|R_{22}|, \tag{3.6}$$

a relation, the same as in the static case,[9] which is a consequence of the time reversal invariance of our potential. The transmission coefficients are related by

$$T_{12}/|H_\lambda^{(1)}(\alpha)|^2=(1+A)T_{21}/|H_\lambda^{(2)}(\beta)|^2. \tag{3.7}$$

In the case under consideration $1+A$ is positive. It can be also checked that, with exception of the case of zero energy, T_{12} and T_{21} have no zeros at physical values of E and λ. So, $\overline{T_{21}}T_{12}$ is strictly positive. From Eqs. (3.3) and (3.4) we have

$$|R_{11}|<1, \quad |R_{22}|<1. \tag{3.8}$$

The first of these relations allows us to obtain, from the definition (2.21),

$$|\rho|=|R_{11}|<1. \tag{3.9}$$

This guarantees the convergence of the Debye expansion.

In order to detect the existence of resonances, we shall examine the modulus of the internal amplitude. From Eq. (2.23), taking account of Eq. (2.16), we have

$$|\mathcal{A}(\lambda,E)|=|1+R_{22}|\,|1-\rho|^{-1}. \tag{3.10}$$

To search for possible maxima of this quantity, we must discuss the behavior of R_{11} and R_{22}, at physical values of λ, as the energy varies from zero to infinity. In the case of S wave, it is immediate to obtain

$$R_{11}(l=0)=-\{A-i\alpha[(1+A)^{1/2}-1]\}/\{A+i\alpha[(1+A)^{1/2}+1]\}, \tag{3.11}$$

$$R_{22}(l=0)=-\{A+i\alpha[(1+A)^{1/2}-1]\}/\{A+i\alpha[(1+A)^{1/2}+1]\}. \tag{3.12}$$

For $l\neq0$, approximate expressions for R_{11} and R_{22} can be obtained in the limits of low energy,

$$R_{11}\simeq-1+i\alpha^{2l+1}2(1+A)^{1/2-l}/A(l+1)[(2l-1)!!]^2, \tag{3.13}$$

$$R_{22}\simeq-1+i\alpha^{2l+1}2/A(l+1)[(2l-1)!!]^2, \tag{3.14}$$

and high energy

$$R_{11}\simeq\{[(1+A)^{1/2}-1]/[(1+A)^{1/2}+1]\}\times\{1+i2(1+A)^{1/2}l!\,\alpha^{-1}\}, \tag{3.15}$$

$$R_{22}\simeq\{[1-(1+A)^{1/2}]/[1+(1+A)^{1/2}]\}\times\{1-i2l!\,\alpha^{-1}\}. \tag{3.16}$$

At intermediate energies, the explicit forms of R_{11} and R_{22} become rather cumbersome. A qualitative behavior of these coefficients can be easily deduced from the corresponding values of $x[1x]$, discussed in the Appendix. In Fig. 1 we show the reflection coefficients for the four lowest angular momentum waves and for two different values of A corresponding to barrier and shallow well, respectively.

In view of the behavior of R_{22}, it is evident that the

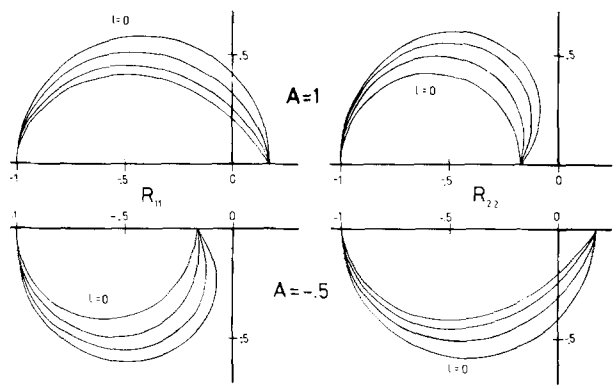


FIG. 1. Internal (R_{11}) and external (R_{22}) coefficients for two different velocity-dependent square potentials of intensities $A=1$ and $A=-0.5$. We have represented the (complex) values of these coefficients in the cases of angular momenta $l=0$, 1, 2, and 3. As the energy varies from zero to infinity, those values describe in the complex plane the trajectories shown in the figure. The trajectories are half-circumferences for $l=0$; for the other values of l they appear consecutively arranged. At zero energy, $R_{11}=R_{22}=-1$; at infinite energy, $R_{11}=-R_{22}=[(1+A)^{1/2}-1]/[(1+A)^{1/2}+1]$.

first factor in the right-hand side of Eq. (3.10) has a smooth dependence on the energy. Therefore, the resonances would occur at energies for which $|1-\rho|$ shows a minimum. In other words, to detect resonances we must look for energies such that

$$|\rho| \simeq 1, \tag{3.17a}$$

$$\arg\rho \simeq 2n\pi, \quad n=0,\pm1,\pm2,\ldots. \tag{3.17b}$$

From Eq. (2.21) we see that

$$|\rho| = |R_{11}|, \tag{3.18a}$$

$$\arg\rho = \arg R_{11} + 2\arg H_\lambda^{(1)}(\beta). \tag{3.18b}$$

The last term in Eq. (3.18b) is an increasing function of β, going from zero to infinity as the energy varies. So, Eq. (3.17b) is fulfilled infinitely many times. In order to discuss Eq. (3.17a), let us distinguish the two cases of $0<A$ (barrier) or $-1<A<0$ (shallow well).

For $0<A$, the requirement $|R_{11}| \simeq 1$ is satisfied at low energies for any A ($R_{11} \simeq -1$) and at high energies for $1 \ll A$ ($R_{11} \simeq +1$). Nevertheless, in both situations it becomes $R_{22} \simeq -1$, that is, $|T_{21}| \simeq 0$. So, the oscillations in $|\mathcal{A}|$ motivated by the factor $|1-\rho|^{-1}$ are damped by the factor $|T_{21}|$. In more physical terms, although the reflection coefficient in the interior region, R_{11}, has the correct value to produce resonance phenomena, these are prevented by the strong direct external reflection of the incoming wave at the surface. The resulting internal amplitude is small and resonances do not appear. The resonancelike maxima of $|\mathcal{A}|$ are the most marked for approximately equal values of the two factors on the right-hand side of Eq. (3.10). This happens for an intensity of the potential $A \simeq 1$, that is, for a barrier of height nearly equal to the energy of the incident particle.

In the case of $-1<A<0$ we have $R_{11} \simeq -1$ at low energies for any A and at high energies for $-1 \lesssim A$. The first situation does not correspond to resonances, since we have $R_{22} \simeq -1$, as in the case discussed in the preceding paragraph. In the second situation, instead, it turns out to be $R_{22} \simeq 1$ and consequently $T_{21} \simeq 2$. So we have, in the case of a velocity-dependent square well of intensity $-1 \lesssim A$, a strong transmission of the incident wave through the potential surface and a big reflection of the interior wave on this surface. Then, all that is needed to obtain a resonance is a constructive interference of the internal waves successively reflected on the surface, that is, the fulfilling of Eq. (3.17b). This condition is satisfied at infinitely many values of the energy for each physical angular momentum. We expect, in this way, an infinite number of resonances for each angular momentum at high energies. These resonances are more and more intense as A approaches -1. In Fig. 2 we have represented, in a Chew—Frautschi plot, the resonances for $A=-0.9$. It can be seen that, for a given Regge trajectory, the resonances become sharper for higher values of the angular momentum.

4. CASE OF A "DEEP" WELL

In the case of a velocity-dependent well of intensity $A<-1$, the internal wavenumber β is pure imaginary

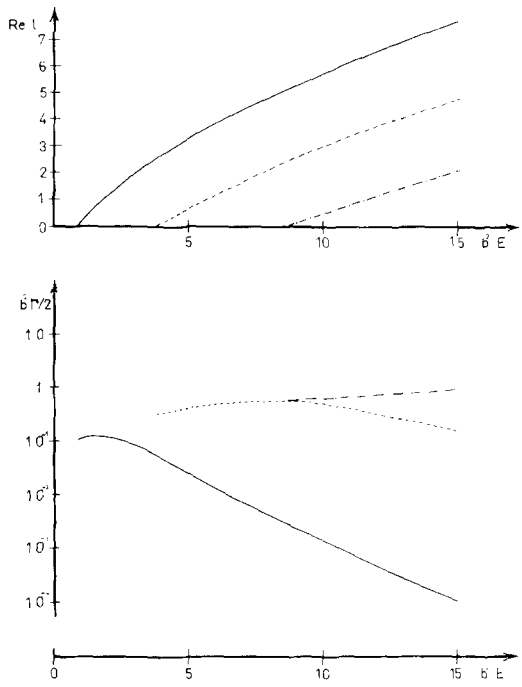


FIG. 2. Chew—Frautschi plot of the Regge trajectories for a velocity-dependent square well potential of intensity $A=-0.9$. There are an infinity of trajectories from which we have represented only the first three. The intersection of each trajectory with the horizontal lines $\mathrm{Re}\,l$ = integer shows the possibility of a resonance at the energy E at which the intersection takes place. In the lower half of the figure we have represented, in adimensional units, the quantity $\Gamma \equiv 2\mathrm{Im}\,l/[d(\mathrm{Re}\,l)/dE]$, which at integer values of $\mathrm{Re}\,l$ gives the width of the eventual resonance. It can be seen that the Regge trajectories connect resonances which become narrower for higher values of the angular momentum.

at positive energies. This corresponds to the fact that, using a terminology borrowed from nuclear physics,[11] the "effective" mass of the particle becomes negative. Equation (2.19) remains still valid for z replaced by α, but not by β. We have, instead,

$$\overline{z[1z]} = z[1z], \quad \overline{z[2z]} = z[2z], \quad \lambda \text{ real}, \ z \text{ imaginary}. \tag{4.1}$$

Using the relations Eq. (2.19) and Eq. (4.1), we obtain from Eqs. (2.17) and (2.18)

$$\mathrm{Im}R_{11} = i\tfrac{1}{2}\overline{T_{12}}T_{21}, \tag{4.2}$$

$$|R_{22}| = 1. \tag{4.3}$$

These two equations replace Eqs. (3.3) and (3.4), which are no more valid in the case under consideration. Equation (4.3) expresses the fact that the incident flux is totally reflected at the surface. There is no transmitted flux since the logarithmic derivative of the internal wavefunction is real. However, the incident external wave originates a stationary internal wave and, conversely, a stationary internal wave gives rise to an outgoing external wave. Hence, we may define reflection and transmission coefficients, in a generalized sense, just as in Sec. 2.

A difficulty arises from the fact that Eq. (3.9) is no more valid in general and, therefore, the Debye ex-

pansion may not be convergent for some values of the energy. This is the same difficulty one encounters in the case of scattering by a static square barrier of a particle with energy below the height of the barrier. In fact, our velocity-dependent well, with $A < -1$, acts as such a barrier for a particle of positive energy. In a forthcoming paper[12] we analyze the convergence of the Debye expansion in a more general context (scattering by a complex potential) which includes the case here considered as a particular one. Nevertheless, the internal amplitude is still given by Eq. (2.23), as it can be easily checked by means of the continuity conditions on the wavefunction and its derivative.

It is well known (see Ref. 10, p. 441) that $H_\lambda^{(2)}(\beta)$, with $\lambda = 1+\frac{1}{2},\ 3+\frac{1}{2},\ 5+\frac{1}{2}, \cdots$, vanishes at a given imaginary value of β. For this reason, we prefer to write Eq. (2.23) in an equivalent form, more suitable to discuss resonances. It is not difficult to obtain

$$\mathcal{A} = 2iF_1/\pi F_2, \tag{4.4}$$

with

$$F_1 = H_\lambda^{(1)}(\beta)/J_\lambda(\beta)\left|H_\lambda^{(1)}(\alpha)\right|^2, \tag{4.5}$$

$$F_2 = \alpha[1\alpha] - (1+A)\beta[\beta] + A/2. \tag{4.6}$$

The factor F_1 is real for physical values of λ. Its modulus shows an exponentially decreasing behavior as the energy goes to infinity. To detect possible resonances we must look for minima of the modulus of F_2. Of course,

$$\mathrm{Im}F_2 = \mathrm{Im}\alpha[1\alpha], \tag{4.7}$$

and, as it can be seen in the Appendix, this imaginary part is always positive and increases monotonously with the energy. The real part,

$$\mathrm{Re}F_2 = \mathrm{Re}\alpha[1\alpha] - (1+A)\beta[\beta] + A/2, \tag{4.8}$$

is also monotonously increasing with the energy, going to infinity at high energies. If it is negative at zero energy, it necessarily vanishes at a certain energy, giving in this way one and only one minimum in the modulus of F_2. The zero energy limit of F_2 follows immediately from that of the Bessel and Hankel functions (see Ref. 10, Eqs. 9.1.7 and 9.1.9). It turns out to be

$$\begin{aligned}\mathrm{Re}F_2(E=0) &= -(l+1/2) - (1+A)(l+1/2) + A/2 \\ &= l(-2-A) - 1.\end{aligned} \tag{4.9}$$

In view of this we can conclude that for $-2 < A < -1$ there will be one resonance of each angular momentum l, and for $A < -2$ there will occur one resonance only in l waves such that $l < (-2-A)^{-1}$. These resonances are very sharp if they appear at low energies. As the energy increases, the exponentially decreasing factor F_1 reduces strongly the modulus of the internal amplitude and the bumps in $|\mathcal{A}|$ occasionated by F_2 cannot be considered as resonances. In the range $-2 < A < -1$, the resonances become sharper as A approaches -2. For $A < -2$, a very sharp resonance of angular momentum l appears if A takes a value slightly above $-2 - 1/l$. For $A = -2 - 1/l$, a zero energy resonance in the l wave occurs. In Figs. 3 and 4 we show the resonances produced by a potential corresponding to $A = -2.2$.

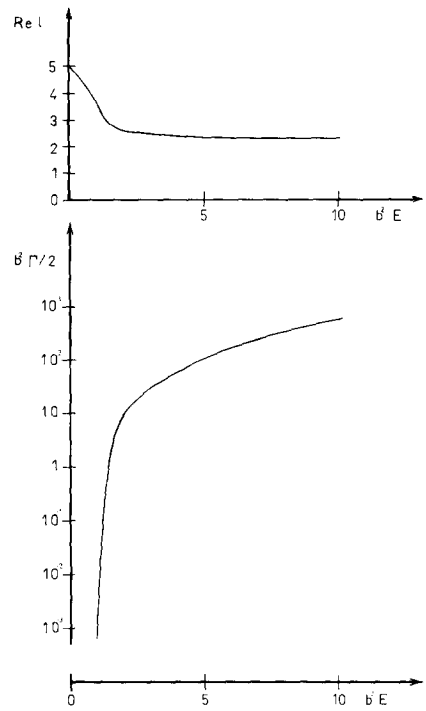


FIG. 3. Chew—Frautschi plot of the Regge trajectories for a velocity-dependent square well of intensity $A = -2.2$. There is only one Regge trajectory related with resonances. It evidences the existence of one resonance of angular momentum $l = 5$ at zero energy, one resonance for $l = 4$ and energy $E \simeq 0.8b^{-2}$, and one resonance for $l = 3$ and $E \simeq 1.2b^{-2}$. The width of these resonances is shown in the lower half of the figure.

ACKNOWLEDGMENTS

We are grateful to the staff of Centro de Cálculo de la Universidad de Zaragoza for the facilities given to us in the use of the computational equipment. It is a pleasure to acknowledge the Ministerio de Educación y Ciencia for a fellowship granted to one of the authors (V.V.)

APPENDIX: LOGARITHMIC DERIVATIVES OF THE HANKEL AND BESSEL FUNCTIONS

In the expression of the reflection and transmission coefficients defined in Sec. 2 enter the logarithmic derivatives of $H_\lambda^{(1)}(z)$, $H_\lambda^{(2)}(z)$, and $J_\lambda(z)$. More precisely, we want to discuss in this Appendix $z[1z]$, $z[2z]$, and $z[z]$, defined by Eqs. (2.6), (2.7), and (2.8), as functions of z.

Let us start with $z[1z]$ in the case of z taking real values, which we shall denote by x. Of course, $x[2x]$ is merely the complex conjugate of $x[1x]$. Let us denote by $M_\lambda(x)$ and $\theta_\lambda(x)$, respectively, the modulus and phase of $H_\lambda^{(1)}(x)$:

$$H_\lambda^{(1)}(x) = M_\lambda(x)\exp[i\theta_\lambda(x)]. \tag{A1}$$

The logarithmic derivative gives

$$[1x] = M_\lambda'(x)/M_\lambda(x) + i\theta_\lambda'(x). \tag{A2}$$

Using the relation (see Ref. 10, Eq. 9.2.21)

$$\theta_\lambda'(x) = 2/\pi x M_\lambda^2(x), \tag{A3}$$

it becomes

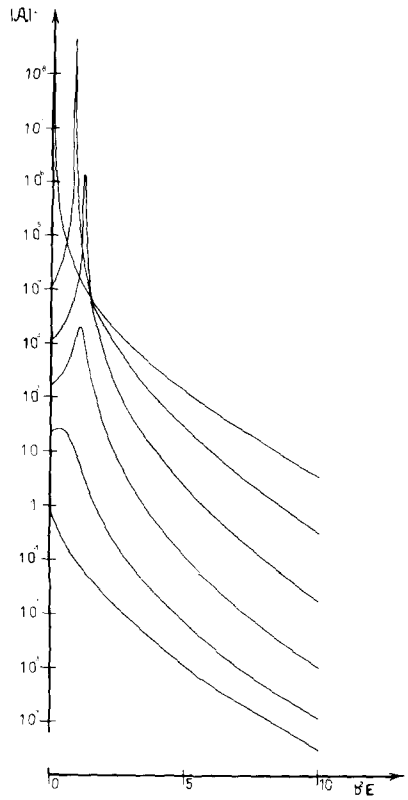


FIG. 4. Internal probability in a velocity-dependent square well of intensity $A=-2.2$. We have represented the square modulus of the internal amplitude [see Eq. (2.23) in the text] for waves of angular momenta $l=0, 1, 2, 3, 4$, and 5. The lowest curve corresponds to $l=0$. The others are consecutively disposed. The zero energy value becomes $|\mathcal{A}|^2=(2l+1)^2|1+A|^{2l+1}/|2l+1+Al|^2$. At infinite energy, all the trajectories tend asymptotically to the line $|\mathcal{A}|^2=4\exp[-2|b^2E/(1+A)|^{1/2}]$. The resonances quoted in the legend of Fig. 3 appear quite clearly as very narrow peaks in the internal probability.

$$x[1x]=x(M_\lambda^2(x))'/2M_\lambda^2(x)+i2/\pi M_\lambda^2(x). \tag{A4}$$

For physical values of λ, the modulus of the Hankel function is given (see Ref. 10, Eq. 10.1.27) by

$$M_{l+1/2}^2(x)=\frac{2}{\pi x}\sum_{k=0}^{l}\frac{(2l-k)!(2l-2k)!(2x)^{2k-2l}}{k![(l-k)!]^2} \tag{A5}$$

and its derivative becomes

$$(M_{l+1/2}^2(x))' \tag{A6}$$
$$=-\frac{4}{\pi x^2}\sum_{k=0}^{l}\frac{(l-k+1/2)(2l-k)!(2l-2k)!(2x)^{2k-2l}}{k![(l-k)!]^2}.$$

It is obvious, from Eq. (A5), that $M_{l+1/2}^2(x)$ is an increasing function of the index l. For a given l, $M_{l+1/2}^2(x)$ decreases monotonously as x increases. From Eq. (A6) we see that $(M_{l+1/2}^2(x))'$ is a negative function whose absolute value decreases as x increases. For a given x, this absolute value is an increasing function of the index l. From these considerations it is not difficult to convince oneself that the behavior of the real and imaginary parts of $x[1x]$ is as shown in Fig. 5. Approximate expressions, valid for limiting values of x, are

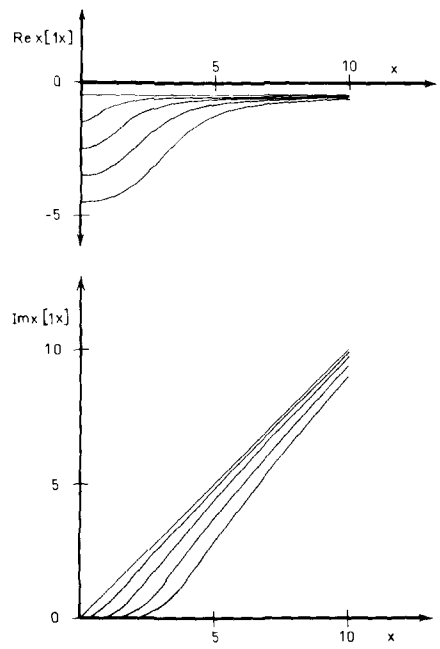


FIG. 5. Real and imaginary parts of the reduced logarithmic derivative of the Hankel function, $x[1x]\equiv xH_\lambda^{(1)\prime}(x)/H_\lambda^{(1)}(x)$, for real argument x and physical values of the index, $\lambda=l+1/2$, $l=0, 1, 2, 3, 4$. The straight lines correspond to $l=0$. The other lines correspond, consecutively, to the remaining values of l.

$$\mathrm{Re}x[1x]\simeq -l-1/2, \quad \text{for } x\to 0, \tag{A7a}$$

$$\mathrm{Im}x[1x]\simeq x^{2l+1}/[(2l-1)!!]^2, \quad \text{for } x\to 0, \tag{A7b}$$

$$\mathrm{Re}x[1x]\simeq -1/2, \quad \text{for } x\to\infty, \tag{A8a}$$

$$\mathrm{Im}x[1x]\simeq x, \quad \text{for } x\to\infty. \tag{A8b}$$

Now, let us discuss briefly the behavior of $z[z]$ for z pure imaginary. Of course, it is a real quantity. If we denote

$$z=iy, \tag{A9}$$

we have the approximate expressions

$$z[z]\simeq l+1/2+y^2/2(l+3/2), \quad \text{for } y\to 0, \tag{A10}$$

$$z[z]\simeq y-1/2+l(l+1)(2y)^{-1}, \quad \text{for } y\to\infty. \tag{A11}$$

At intermediate values of y, the behavior of $z[z]$ is that shown in Fig. 6.

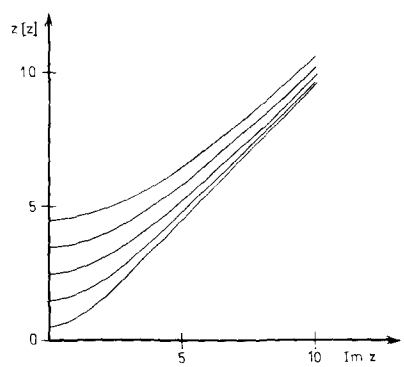


FIG. 6. Reduced logarithmic derivative of the Bessel function $z[z]\equiv zJ_\lambda'(z)/J_\lambda(z)$, for pure imaginary argument z and physical values of the index, $\lambda=l+1/2$, $l=0, 1, 2, 3, 4$. The curves are ordered from bottom to top with increasing l.

[1]K. M. McVoy, L. Heller, and M. Bolsterli, Rev. Mod. Phys. **39**, 245 (1967).
[2]J. E. Beam, Am. J. Phys. **38**, 1395 (1970); H. C. Ohanian and C. G. Ginsburg, Am. J. Phys. **42**, 310 (1974); H. D. Meyer and K. T. Tang, Z. Physik A **279**, 349 (1976).
[3]F. Henyey, Am. J. Phys. **39**, 1477 (1971).
[4]H. M. Nussenzveig, Phys. Rev. D **6**, 1534 (1972); D. Bollé and T. A. Osborn, Phys. Rev. D **11**, 3417 (1975).
[5]R. M. More and E. Gerjuoy, Phys. Rev. A **7**, 1288 (1973); G. Garcia Calderon, Nucl. Phys. A **261**, 130 (1976); G. Garcia Calderon and R. Peierls, Nucl. Phys. A **265**, 443 (1976).
[6]E. M. Ferreira, N. Guillen, and J. Sesma, J. Math. Phys. **8**, 2243 (1967); **9**, 1210 (1968).
[7]E. M. Ferreira, R. Merí, and J. Sesma, J. Math. Phys. **18**, 944 (1977).
[8]M. Razavy, G. Field, and J. S. Levinger, Phys. Rev. **125**, 269 (1962).
[9]H. M. Nussenzveig, J. Math. Phys. **10**, 82 (1969).
[10]M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions* (Dover, New York, 1965).
[11]W. F. Hornyak, *Nuclear Structure* (Academic, New York, 1975), p. 219.
[12]A. Cruz and J. Sesma (to be published).

Null infinity is not a good initial-data surface[a)]

Robert Geroch

Enrico Fermi Institute, Chicago, Illinois 60637
(Received 28 November 1977)

An example is given of a space-time which is asymptotically flat and globally well-behaved, and yet which admits a nonzero Maxwell field, with zero source, having no incoming radiation from past null infinity.

1. INTRODUCTION

There are a number of circumstances in general relativity in which it would be of interest to know whether or not a zero-mass field without sources in a well-behaved space–time is uniquely determined by its asymptotic behavior at past null infinity, i.e., by its incoming radiation. For example, quantum field theory in curved space–times is usually studied[1] in an S operator framework. But this framework requires that in and out vacuum states can be identified, which requires in turn in and out creation and destruction operators, which requires that the classical fields in the space–time can be decomposed into their asymptotic positive and negative frequency parts, which requires, finally, that these classical fields can be characterized by their asymptotic behavior. Suppose, however, that one's space–time admits nonzero fields with zero incoming radiation. Then the resulting "classical particle creation," uncontrollable from null infinity, would, among other things, make ambiguous the S operator. In the study[2] of the stability of isolated gravitating systems, to take a second example, one sometimes considers perturbations which arise from incident radiation from the distant past. Suppose, however, that additional perturbation fields could appear "spontaneously," without their having been recorded as incident radiation. Then there would arise the possibility that one's system is unstable to only these perturbations, and thus is physically unstable, although an analysis in terms of incident radiation would not detect this fact.

We shall here be concerned with the status of the following

Statement: Consider a space–time which is asymptotically flat at past null infinity, and which satisfies certain other global conditions. Then any zero-source, zero-mass field in that space–time, which vanishes sufficiently quickly in the limit at past null infinity, must vanish.

This statement is true, for example, for Minkowski space–time. Indeed, the zero-mass fields in this case satisfy Huygen's principle, i.e., the Green's function has support on the null cone. Using the Green's function, the field at any point of Minkowski space–time can be written as an integral over the asymptotic field at past null infinity. Hence, whenever the field vanishes sufficiently quickly asymptotically, it must vanish everywhere. One might have thought that some sort of similar argument would work also in the presence of curvature. Although curvature, of course, in general destroys Huygen's principle, "virtually all" of the contributions to the field at a point of the space–time, via a Green's function, should come from initial data at null infinity, for the surface at past null infinity is "practically a Cauchy surface, lacking only the single point at past timelike infinity." It would seem to be difficult to squeeze, via a zero-mass field, any real information into the space–time through this "single point."

There is a simple example which shows that no proof of the statement above can use only elementary properties of hyperbolic equations. We claim. There exists a smooth potential V in Minkowski space–time, strictly vanishing in a neighborhood of past null infinity, together with a nonzero φ satisfying $\nabla^2\varphi = V\varphi$ (∇^2 = wave-operator) and vanishing as quickly as one wishes in the limit at past null infinity. First recall that [in the usual coordinates, with $r = (x^2 + y^2 + z^2)^{1/2}$ and $u = t - r$], $f(u)/r$, for any smooth function f, is a solution of the wave equation, smooth for $r > 0$. Set, for $r \geq 1$, $\varphi = f(u)/r$, with f any smooth positive function approaching zero quickly as $u \to -\infty$ (i.e., in the limit at past null infinity). For $r < 1$, let φ be any smooth positive function such that the join at $r = 1$ is smooth. Now set $V = \nabla^2\varphi/\varphi$, so this V of course has all the required properties. This example, however, is not conclusive with respect to the statement above, for one could think of V as endowing φ with an "effective mass."

The present result is that, in the Maxwell case under what seem to be rather strong "other global conditions," the statement above is false.

We turn now to the question of what is an appropriate set of global conditions. That some conditions will be required is illustrated by the following example. Consider the time reverse (in order to deal with past null infinity) of a solution for a collapsing, spherically symmetric, dust cloud, Schwarzschild outside. This space–time possesses a Cauchy surface which enters the (now time-reversed) black hole. Choose on such a surface data for Maxwell's equations which are nonzero inside the hole and zero outside, and evolve to obtain a Maxwell field on the entire space–time. Since no timelike or null curve, beginning inside the black hole, can reach past null infinity, the resulting Maxwell field, while nonzero, will even vanish in a neighborhood of past null infinity. The physical reason for this example is of course that there are regions of the space–time which are unable to communicate with past null infinity. One could eliminate examples of this type

[a)]Supported in part by the National Science Foundation, under contract PHY 76-81102.

 0022-2488/78/1906-1300$1.00

by demanding that such communication be possible, i.e., that the future of past null infinity be the entire space–time. But it turns out that this condition alone is not enough. Consider, in Minkowski space–time, the Maxwell field which results from an oscillating point dipole on the t axis, where the amplitude and frequency of the oscillations die off quickly into the past. Now excise the t axis from the space–time. There results a space–time which satisfies the communication condition, and yet which supports a smooth nonzero Maxwell field with no incoming radiation from past null infinity. The problem now is that one has introduced a singularity which is capable of "emitting electromagnetic radiation at all times," thus giving rise to fields on the space–time not recorded at past null infinity. One could eliminate such examples by demanding that one's space–time admit a Cauchy surface[3]—i.e., that there be *some* initial-data surface for fields on the space–time. This example, of course, has no Cauchy surface.

We conclude, then, that a possibe set of "other global conditions" is the existence of a Cauchy surface and that the future of past null infinity be the entire space–time. There exists, however, another global condition, which on the one hand is even stronger than these two taken together, and, on the other, has a physical basis. One might demand instead that every maximally extended, past-direct null geodesic reach past null infinity.[4] One thinks of one's zero-mass field, in the optical limit, as yielding "particles" which travel along null geodesics, and so this condition requires that all such particles reach the asymptotic region. Our statement with this global condition, then, is essentially that if what happens internally is determined by what happens asymptotically in the optical limit, then it is also determined prior to that limit.

Our example will satisfy all of these conditions.

2. THE EXAMPLE

We shall obtain an example of a space–time which is Minkowskian outside of a spatially bounded (defined with respect to the external metric) timelike world tube. Every null geodesic will emerge from, and remain outside of, that world tube. Thus, the space–time will be topologically $\mathbb{R}^4$, will admit a Cauchy surface, will have no causality violations, will admit well-behave spacelike slices, and will be asymptotically flat and even asymptotically simple.[4] It will also be spherically symmetric. This space–time will carry a smooth nonzero Maxwell field, with zero source, which vanishes exponentially at past null infinity. (That is, for some positive r_0 the Minkowskian components of this field, multiplied by $\exp(r/r_0)$, will vanish in the limit along every past-directed null geodesic.) In particular, this field will be asymptotically regular at past null infinity, with zero data there. The Maxwell field will also have zero angular momentum. Since Maxwell's equations are conformally invariant, one can by means of a conformal rescaling also have the space–time nonsingular, i.e., timelike and null complete. We do not impose Einstein's equation.

Let M, g_{ab} be spherically symmetric, i.e., admit three independent Killing fields whose commutation relations are those of the rotation group and whose integral surfaces are (except, of course for the origin) metric 2-spheres. Define the usual radial function $\hat{r}$—such that $2\hat{r}^2$ is the sum of the squares of the norms of the three Killing fields—and denote by ϵ_{ab} the alternating tensor of the 2-spheres. Let F_{ab} be a zero-source Maxwell field on this space–time, and set $\varphi = \hat{r}\epsilon^{ab}F_{ab}$. Then, applying the wave operator, using Maxwell's equations and spherical symmetry, one sees that φ satisfies the wave equation. Denote by S the manifold of 2-spheres of spherical symmetry. (For example, this manifold for the Schwarzschild solution, is the Penrose diagram.) Then this S acquires a metric h_{ab} of signature $(-, +)$, and also a function $\hat{r}$. Now consider the special case in which $\varphi = \alpha\hat{r}^{-1}u$, where α is a spherical harmonic, principal quantum number l, on the 2-spheres, and u is a function on S. Then the wave equation becomes an equation for u on the two-dimensional manifold S,

$$D^2u = \hat{r}^{-2}l(l+1)u, \tag{1}$$

where D_a is the derivative operator of S, h_{ab}, and $D^2 = h^{ab}D_aD_b$.

The converse of all this is the following. Let S be a two-dimensional manifold with metric h_{ab} of signature $(-, +)$, nonnegative scalar field $\hat{r}$, and scalar field u satisfying (1) for some positive integer l. Then one obtains from S, h_{ab}, and $\hat{r}$ a spherically symmetric space–time (with no singularity at the origin, provided h_{ab} and $\hat{r}$ have proper behavior as $\hat{r} \to 0$), and from u a Maxwell solution, via the formula

$$F_{ab} = \tfrac{1}{2}\alpha\hat{r}^2u\epsilon_{ab} - [l(l+1)]^{-1}\nabla_m\alpha\epsilon_{[a}{}^m\nabla_{b]}u, \tag{2}$$

where α is a spherical harmonic, principal quantum number l, on the 2-spheres, and ∇_a denotes the gradient in this space–time.

The idea is the following. Let S, h_{ab}, and r be the manifold, metric, and field which results from this construction in flat space–time, so h_{ab} is flat and D_ar is a constant, unit spacelike vector field on S. We look for a function u which satisfies, instead of (1),

$$D^2u = r^{-2}l(l+1)u + f, \tag{3}$$

where f, the source, is some function on S. We now try to arrange matters so that f is nonnegative, f vanishes except between two finite r values (so, in particular, f vanishes near past null infinity and near the origin), and such that u is positive in the support of f. Having done this, we set

$$\hat{r}^{-2} = r^{-2} + [l(l+1)]^{-1}u^{-1}f, \tag{4}$$

to obtain our solution, h_{ab}, u, $\hat{r}$, of (1). Finally, one must check that, in the resulting space–time (defined by S, h_{ab}, $\hat{r}$), every null geodesic escapes to null infinity. The idea, in short, is to "solve" Maxwell's equations for the metric rather than for the Maxwell field.

Fix once and for all the positive integer l. We first consider the retarded Green's function for (3), i.e., the retarded solution u for f a δ function located, say, at $r=1$, $t=0$. It is easy to write out this Green's func-

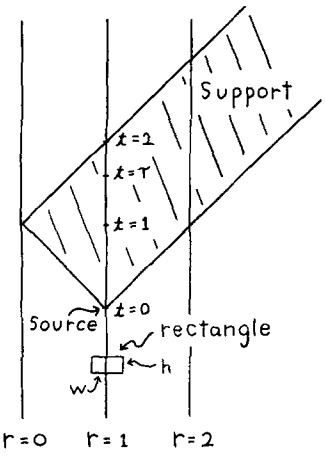


FIG. 1. The support of the retarded Green's function, in the rt plane, for a source at $r=1$, $t=0$. Also shown is one of the rectangles, of width w and centered on the line $r=1$, which is required for the example.

tion explicitly, since (3) is essentially the wave equation with source in Minkowski space–time, while one knows the Green's function for the wave equation, δ(retarded time)/(spatial distance). The resulting function u, we claim, has support the half-infinite rectangle shown in Fig. 1 (an immediate consequence of Huygen's principle for the wave equation), is smooth in the interior of its support, is bounded, and is positive at $r=1$, $t=\tau$, for some $1<\tau<2$. It follows from this last property that there exists positive w sufficiently small that this u is positive even in the square of side $2w$ centered at $r=1$, $t=\tau$, and in addition $1+4w<\tau<2-4w$.

Fix once and for all a smooth, nonzero, nonnegative function A of two variables, such that $A(r,x)$ vanishes unless $|r-1|<w/2$ and $|x|<w/2$. We shall be considering various rectangles in the rt plane, all having width w, but with various heights; and all centered at $r=1$, but at various t values. By a source of strength a in the rectangle of height h centered at $t=t_0$, we mean $f(r,t)=aA(r,wh^{-1}(t-t_0))$; whose support of course is this rectangle. We now claim that there exists a constant c sufficiently large that the retarded solution u of (3) which results from a source of strength $a>0$ in the rectangle of height $h\leq w$ at $t=t_0$ has the following properties: (i) In the rectangle of height h centered at $t=t_0$, $|u|\leq cah^2$. (ii) In the rectangle of height w centered at $t=t_0+\tau$, $u>c^{-1}ah$. (iii) In the rectangles of height w centered at $t=t_0+n\tau$, for $n\neq 0$, 1, u vanishes. (iv) Everywhere, $|u|\leq cah$. Indeed, condition (i) follows from boundedness of the Green's function and its support. (The region over which the integration must be performed has area bounded by h^2.) Condition (ii) follows from the first defining property of w. (The integral of the source over its support is bounded by ah.) Condition (iii) follows from the second defining property of w, and the support of the Green's function. Finally, condition (iv) follows again from boundedness of the Green's function.

We now construct our example. Fix an infinite sequence of rectangles, at $t=0$, $-\tau$, -2τ, $\cdots$, of heights $h_1, h_2, \cdots$ (each less than w), and in which there are placed sources of positive strengths $a_1, a_2, \cdots$. Let the sum of these sources be the f in (3), and compute from this equation the resulting retarded u. (The integral for u must converge, for the support of the Green's function is such that no more than two rectangles will contribute to the u at any one point.) We must choose these h's and a's such that three conditions are satisfied: That u is positive in the support of f [in order that (4) will define a smooth $\hat{r}$], that our Maxwell field decays exponentially at past null infinity, and that every null geodesic escapes to null infinity. Clearly, the simultaneous satisfaction of these conditions will complete the example.

The first two conditions are easy. By (iii) above, the only contributions to u in the ith rectangle come from the sources in the ith and the $(i+1)$st. But these contributions are bounded, respectively, by properties (i) and (ii) above. Thus, we may ensure that u is positive in the ith rectangle by demanding $c^{-1}a_{i+1}h_{i+1}>2ca_ih_i^2$. (We have here inserted an extra factor of two on the right in order to have available, for later use, a lower bound on the value of u in the ith rectangle, namely $\frac{1}{2}c^{-1}a_{i+1}h_{i+1}$.) Furthermore, by (iv) above, our Maxwell field will decay exponentially at past null infinity provided a_ih_i is bounded by $\exp(-i/i_0)$ for some positive i_0.

We must finally impose the third condition, that all null geodesics escape. Consider a null geodesic, with tangent vector l^a, in our (curved, with $\hat{r}$) space–time. Denote by L the total (conserved) angular momentum of this geodesic, i.e., L^2 is the sum of the squares of the inner products of l^a and the three Killing fields. Projecting this geodesic into S, we obtain a curve in S (still parametrized by the original affine parameter) whose tangent vector we denote v^a. Then nullness of l^a becomes $v^av_a=-L^2\hat{r}^{-2}$. The geodesic equation in space–time, projected into S, yields the equation for this curve in S,

$$v^mD_mv^a=-\tfrac{1}{2}L^2D^a(\hat{r}^{-2}). \tag{5}$$

We must arrange that all curves satisfying (5) escape to infinity. We first note that, outside of the rectangles, $\hat{r}=r$, so (5) is there the equation in S for null straight lines in Minkowski space–time, while of course all such curves do escape. Failure to escape, therefore, must be due to deflection of these curves by the rectangles, in which $\hat{r}\neq r$. By the choice of τ and w, a past-directed null geodesic in Minkowski space–time, emerging from the ith rectangle, can meet no other rectangles except possibly for the $(i+1)$st. Thus, the only past-directed null geodesics in the curved space–time which are even candidates for being trapped are those which meet all the rectangles beyond a certain one. The deflection suffered by a null geodesic on passing through a rectangle is estimated from (5). Contracting this equation with D_at, replacing v^mD_m by $(v^mD_mt)\,d/dl$, we obtain

$$d/dt(\log(-v^mD_mt))=-[L^2(v^mD_mt)^{-2}][D_atD^a(\hat{r}^{-2})]. \tag{6}$$

The first factor on the right is bounded by $\hat{r}^2$, and hence, from (4), is bounded. The second factor on the right is, by (4), a multiple of $D_atD^a(fu^{-1})=u^{-1}D_atD^af-u^{-2}fD_atD^au$. But, for the ith rectangle, the two terms on the right are bounded, respectively, by multiples of

$(a_{i+1}h_{i+1})^{-1}(a_ih_i^{-1})$ and $(a_{i+1}h_{i+1})^{-2}(a_i)(a_{i+1}h_{i+1})$. Hence, the right side of (6) is bounded by a multiple of $a_ih_i^{-1}a_{i+1}^{-1} \times h_{i+1}^{-1}$. Integrating (6) along the geodesic through the ith rectangle, we see that the fractional change in v^mD_ml is bounded by a multiple of $a_ia_{i+1}^{-1}h_{i+1}^{-1}$. This, then, is an upper bound on the deflection (as measured by the fractional change in v^mD_ml) of a null geodesic on passing through the ith rectangle. If this deflection is sufficiently small, then a null geodesic reaching the ith rectangle from the $(i-1)$st will not be able to reach the $(i+1)$st, and so will escape. We conclude, then, that all null geodesics will escape provided that, for any n, however large, and for any positive K, we have $a_ia_{i+1}^{-1}h_{i+1}^{-1} < K$ for some $i > n$.

Our example will be completed, then, by displaying a sequence $h_1, h_2, \cdots$ of positive numbers all less than w and a sequence $a_1, a_2, \cdots$ of positive numbers, such that $c^{-1}a_{i+1}h_{i+1} > 2ca_ih_i^2$ for all i, that a_ih_i is bounded by $\exp(-i/i_0)$ for some positive i_0, and that $a_ia_{i+1}^{-1}h_{i+1}^{-1}$ is not bounded away from zero. But suitable sequences are the following. Fix sufficiently small positive ϵ, and set, for i odd, $h_i = \epsilon^i$, $a_i = \epsilon^{2i}$, and, for i even, $h_i = \epsilon^{6i}$, $a_i = \epsilon^{-5i}$.

3. CONCLUSION

One could ask whether or not similar examples would be expected for fields other than the electromagnetic. Of course, the question is of interest only for fields or systems of fields whose equations admit an initial-value formulation. Otherwise, e.g., for Maxwell fields with freely specifiable current, one has available a source for the field other than the asymptotic region, and so examples are easy to write down. Inspection of the present example reveals that it is also an example for the zero-mass Klein–Gordon equation ($\nabla^2\varphi = 0$). One would perhaps also expect similar examples for the zero-mass, conformally invariant Klein–Gordon equation $[(\nabla^2 - \frac{1}{6}R)\varphi = 0]$, and for the neutrino equation. For the nonzero mass fields, there are presumably examples even simpler than the present one. These fields, in the optical limit, yield "particles traveling along timelike geodesics," while such geodesics can never reach past null infinity. Examples for various nonlinear systems of coupled fields, e.g., charged Klein–Gordon and Maxwell, should also be simpler, for such a system could be regarded as having, by virtue of the coupling, an effective mass. Finally, one might ask the question for linearized Einstein fields. This case appears to be considerably more difficult. First, one must contend with the freedom of gauge transformations. The choice of gauge is important and delicate, for a poor choice can[5] result in perturbation fields whose asymptotic behavior does not all reflect the "true" behavior of the asymptotic gravitational field. Second, one must find a background space–time in which the linearized fields are to be studied. This background, say for consideration of the gravitational field alone, would have to be nonflat (for there are no counterexamples for perturbations off Minkowski space–time), asymptotically simple, and satisfying Einstein's equation with zero source. But no such space–time are known.

There is of course a second aspect of this issue, in which one deals with existence rather than uniqueness. By "existence," we mean in the stronger sense: Does every set of initial data at past null infinity give rise to some solution? Although there seems to be no solid evidence on this question, one's suspicion might be that existence will also fail. Unfortunately, it appears that counterexamples will be more difficult to obtain, for, while uniqueness requires only the display of a single field, existence requires consideration of many possible fields.

The important question, however, is that of how common and of what character are the space–times which do have the property that the asymptotic behavior of zero-mass fields suffices to determine the field in the entire space–time. One would of course like some positive results—to the effect that many space–times do have this property. An example of such a result—although far too weak to provide any real evidence on this issue—is the following. In a stationary, asymptotically simple space–time, a stationary Maxwell field which vanishes sufficiently quickly at infinity must vanish. Let S be an asymptotically flat spacelike slice in the space–time, with unit normal n^a. Set, at points of S, $D_a = F_{ab}n^b$, and let E_a be the orthogonal projection into S of $F_{ab}\xi^b$, where ξ^b is the Killing field. Then each of D_a and E_a is a vector field on S; the former with vanishing divergence and the latter with vanishing curl. So, the integral of D_aE^a over S vanishes. Similarly for the magnetic fields. So, the sum of these two integrals—which is the total energy of the electromagnetic field with respect to the Killing field—must vanish. But this is compatible with a nonzero Maxwell field only if the Killing field ξ^a becomes spacelike in some region. But now a null geodesic, beginning in this region with negative ξ energy, will be unable to ever reach the asymptotic region, violating asymptotic simplicity.

One could imagine two approaches to stronger positive results. The first begins with the observation that the present set of global conditions on our space–times were, after all, chosen merely to eliminate a few obvious physical situations under which fields could enter the space–time without their having been recorded asymptotically. Perhaps some further, physically reasonable, global conditions would eliminate the present examples. It is not easy, however, to see what such conditions should be, for this example is quite tame globally. The second approach would be to accept the present set of global conditions, and instead ask whether space–times in which fields *are* determined by their asymptotic behavior are "generic" in any sense. Perhaps, for instance, the present example would be destroyed by a small change of the space–time metric. Is there some simple perturbation calculation which would decide this?

[1]See, for example, S.W. Hawking, Commun. Math. Phys. **43**, 199 (1975); R. Wald, Commun. Math. Phys. **45**, 9 (1975).

[2]See, for example, K. Thorne, Appl. J. **158**, 1 (1969); C.V. Vishveshwara, Phys. Rev. D **1**, 2870 (1970).

[3]See, e.g., S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U.P., London, 1973).

[4]R. Penrose, Proc. Roy. Soc. A **270**, 193 (1962).

[5]See, for example, R. Geroch and B. Xanthopoulos, J. Math. Phys. **19**, 714 (1978).

A particle model based on stringlike solitons

U. Enz

Philips Research Laboratories, Eindhoven, The Netherlands
(Received 5 December 1977)

This paper deals with two scalar fields $\theta(x,y,z,t)$ and $\phi(x,y,z,t)$ which are governed by two coupled nonlinear differential equations. Some of the spatial field distributions of θ and ϕ are topologically stable and represent solitons in three-dimensional space. The simplest stable solitons are identified with the electron and the positron. The asymptotic solutions of the fields are studied. θ is shown to fall off asymptotically as $\pm a/r$, where a is a constant related to the elementary charge and r the distance from the "site" of the soliton.

A recent paper[1] describes a new class of solitons which are represented by field configurations of two scalar fields $\theta(x,y,z,t)$ and $\phi(x,y,z,t)$. Some of these solitons are topologically stable in three-dimensional space. The field configurations, which are free from singularities, are the stationary solutions of a simple least-action principle. An attempt has been made to identify these solitons with elementary particles. The aim of the present paper is to clarify the asymptotic properties of the fields and to give additional agruments for such an identification. It will be shown that for a specific form of the basic action principle θ falls off as $\pm a/r$ asymptotically, where r is the distance from the "site" of the soliton and a is a constant, which is related to the elementary charge. The least action principle is assumed to be

$$\delta W = 0, \quad W \int w d^3x\, dt \tag{1}$$

and is based on a density

$$w = K \sin^2\theta + AD_0 + ED_0^s, \tag{2}$$

where the abbreviation

$$D_0 = (\nabla\theta)^2 - c^{-2}\theta_t^2 + [(\nabla\phi)^2 - c^{-2}\phi_t^2]\sin^2\theta \tag{3}$$

is introduced. The *energy* density is defined as an expression analogous to (2) in which the time like derivatives enter with the positive sign. The scalar fields θ and ϕ have the character of angular variables in an Euclidean space of three dimensions u_i, where they determine the direction of a unit vector $\mathbf{n}$. The dimensions of the constants A, K, and E follow from (1) and (2). A term ED_0^s, where $s \geq 2$, is necessary to assure the stability of the soliton in three-dimensional space.[1] Equation (3) demonstrates the Lorentz invariance of the model. The Euler equations corresponding to the least action principle (1) are two coupled nonlinear partial differential equations of second order for θ and ϕ.[1] They are not reproduced here but for the special case $K=0$ [see Eqs. (7), (8)]. In the general case, no solution of the Euler equations[1] was obtained However, in the case of cylindrical symmetry, with coordinates ρ, φ, z, solutions were found by assuming $\theta = \theta(\rho)$ and $\phi = \varphi + kz$, where k is a constant. Equation (3) then takes the form

$$D_0 = \theta_\rho^2 + (\rho^{-2} + k^2)\sin^2\theta. \tag{4}$$

From Eq. (4) it can be seen that singularity-free solutions at $\rho = 0$ imply $\sin\theta(0) = 0$ or $\theta(0) = M\pi$, where M is an integer. We impose the further boundary condition $\theta = 0$ for $\rho \to \infty$. Solutions based on (4) represent field configurations which are infinitely extended in the z direction. The field θ is equal to $M\pi$ on the entire z axis and falls to zero for $\rho \to \infty$. These solutions are stable in a topological sense, related to the fact that ϕ changes by 2π on any simple closed path containing the z axis.

The density w contributes only within a cylindrical region of finite lateral dimension appreciably to the integral (1). Such structures may be visualized as *strings* having an energy proportional to their length and thus a constant "tension." The line defined by $\theta = M\pi$, the z axis, is called the *core* of the string. Two types of strings with different lateral dimensions are now considered, both of them based on Eq. (4).

(i) K, E and A are all assumed to be different from zero. A static solution $\theta(\rho)$ was found numerically for $s=2$ and $k=0$.[1] The function $\theta(\rho)$ decreases exponentially to zero for $\rho \to \infty$ within some multiples of the length $l_0 = (E/K)^{1/4}$. For $M=1$, the energy per cm of the string was calculated to be $2\pi\,(EK)^{1/2} I_1(\gamma)$, where I_1 is a dimensionless integral depending on $\gamma = EK/A^2$. For $\gamma = 1$, this integral turned out to be about 12.9.

(ii) If we assume $E \neq 0$ and $K = 0$, no solution with finite l_0 exists for $k = 0$. However, for $k \neq 0$ solutions with finite lateral dimensions *do* exist. These solutions represent *twisted strings*. Their lateral "size" is of the order of $l^{1/2}k^{-1/2}$, where the length

$$l = (E/A)^{1/2} \tag{5}$$

as introduced. In this paper, mainly strings of type (ii) are considered: the Euler equations are given here for this case. Using the basic action principle (1) in dimensionless form

$$\delta W = 0, \quad W = (E/c)\int (D + D^2)\, d^4\xi, \tag{6}$$

where $D = l^2 D_0$ and the ξ's are dimensionless coordinates expressed in units of l, the Euler equations read

$$2(1+2D)\Box\theta + 4(\nabla\theta\nabla D - \theta_\tau D_\tau) - (1+2D)[(\nabla\phi)^2 - \phi_\tau^2]\sin 2\theta = 0 \tag{7}$$

and

$$[(1+2D)\Box\phi + 2(\nabla\phi\nabla D - \phi_\tau D_\tau)]\sin\theta + 2[(1+2D)(\nabla\phi\nabla\theta - \phi_\tau\theta_\tau)]\cos\theta = 0, \tag{8}$$

where $\tau \equiv \xi_4$.

 0022-2488/78/1906-1304$1.00

We now consider field configurations having the property that the major part of the total field energy is found inside a sphere of finite radius. Field configurations of such a type are constructed by forming *closed twisted strings*, i.e., strings whose core forms a closed curve. It is assumed that they represent approximative solutions to (1). The field remains continuous if the closed string contains N 2π twists on its circumference, where N is an integer. It can be shown[1] by topological arguments that strings of type (i) and (ii) with $s \geq 2$ are stable if $N=1$, but unstable if $N=0$. For $N=1$ a minimum of the total action (6) occurs at a finite circumference $2\pi R_0$ of the string. For $N \geq 2$, the stability problem is more complicated and cannot be decided by topological arguments. The minimum energy and the corresponding stable string radius R_0 of a closed static string of type (i) have been calculated with the aid of a Ritz approximation[1] for $N=1$ and $s=2$.

The continuity of ϕ furthermore implies the existence of either an open ended, infinite core $\theta=0$ or a closed, finite core $\theta=0$ interlocking with the string $\theta=\pi$, the two configurations being denoted by $C=0$ and $C=1$, respectively. The structure $C=1$ is symmetric with respect to $\theta=\pi/2$, in accordance with the symmetry of the basic expression (2) of the energy density.

A further integer P may be introduced, describing the sense of rotation of the twist. P takes the value $+1$ for a right-handed twist and the value -1 for a left-handed twist, thus describing a mirror symmetry between the two structures. Summarizing, the symmetry and topological properties of closed strings of type (i) and (ii) are characterized by the integers N, M, C, P.[2] Each of these structures has a finite size and a finite energy. These structures are solitons having a "site" or location in three dimensional space. The main part of the total field energy is concentrated about this site.

We now discuss *the asymptotic properties* of the fields θ and ϕ, making the restriction to the simplest closed string of type (ii) with $N=M=C=1$, $P=\pm 1$. The soliton is assumed to be at rest and to be located at the origine $r=0$ of a polar coordinate system. Static asymptotic solutions of spherical symmetry with the boundary conditions $\theta(r) \to 0$ and $\phi \to$ const for $r \to \infty$ are considered. It is important to realize, that the boundary condition for ϕ is compatible only with solutions described by $C=1$ and not with those with $C=0$. This property follows directly from the topology of the structures $C=0$ and $C=1$. Equation (8) is fulfilled by $\phi=$ const for any θ. Equation (3) then reduces to

$$D=\theta_r^2 \tag{9}$$

if r is measured in units of l. With (9) Eq. (7) takes the form

$$(1+2\theta_r^2)\left(\theta_{rr}+\frac{2}{r}\theta_r\right)+4\theta_r^2\theta_{rr}=0. \tag{10}$$

Equation (10) separates and has the solution

$$\theta+\theta_\infty=\int_0^q p\frac{dr(p)}{dp}\,dp, \tag{11}$$

where

$$r(q)=a^{1/2}|q|^{-1/2}(1+2q^2)^{-1/2} \tag{12}$$

and $q=\theta_r$. The complete asymptotic solution of (10) which contains the integration constants θ_∞ and $a>0$, is found from (11) and (12) by elimination of q. The boundary conditions imply $\theta_\infty=0$. As (10) applies only to the case $r \gg 1$, we expand $\theta(r)$ in this limit to yield

$$\theta(r) \approx \pm a/r. \tag{13}$$

The value of a is not determined by the asymptotic solution alone. A preliminary analysis, however, seem to show that the complete Euler equation (7), (8) determine a unique value of a for all structures $C=1$. For $C=0$ we find $a=0$, $\theta(r)$ then decays faster than $1/r$. In conclusion we find that closed strings of type (ii) with $C=1$ have an asymptotic solution $\pm a/r$. It can be shown that an asymptotic $1/r$ dependence of θ also results for strings of type (i), but only for a time dependence of ϕ of the form ωt, with $\omega=c(K/A)^{1/2}$, and only if $s \geq 3$, and not for $s=2$ as has been incorrectly stated in Ref. 1.

A physical interpretation is now given. The idea is to equate the total energy of a given field structure to the rest energy of a particle and in this way to establish a correspondence between solitons and particles. In particular, by hypothesis, the solitons characterized by $N=M=C=1$, $P=\pm 1$ are identified with the electron and the positron. This assignment is the same as made before,[1] but is now based on strings of type (ii). An argument for such a choice is that these solitons are the simplest stable structures having an asymptotic behaviour $\pm a/r$. The total energy is defined as the spatial integral I_1 over the energy density, following in dimensionless form from (6) (timelike derivatives with positive sign!). We then put

$$mc^2=(E/l)I_1, \tag{14}$$

where m is the mass of the electron. The fundamental length l is determined by identifying the circumference $2\pi R_0$ of the stable string with the Compton wavelength:

$$R_0=\hbar/mc. \tag{15}$$

The stable string radius R_0 follows in principle from (6) in units of l, so that l is determined by (15). The motivation for the assignment (15) is that the Compton wavelength can be regarded as a structural element of the electron.[3,4] A further argument concerns the corresponding frequency $mc^2/\hbar$ obtained if the twist is thought to move along the string with velocity c, thus establishing an internal motion or spin. At a given point in space, θ shows then a timelike variation with the above frequency. We think that this timelike variation of θ if a property of the solutions of the basic action principle. The whole structure can then be considered as an extended oscillator of frequency $mc^2/\hbar$. This concept makes a link between the proposed structure and quantum mechanics, which was originally[5] based on moving oscillators of this frequency. Equations (14) and (15) determine both available constants E and l, so that no further adjustment is possible. The constants E, l, and c are therefore the basic physical constants of this model.

If, as in Ref. 1, the energy density in the asymptotic part of the solition (here $Al^{-2}a^2r^{-4}$) is equated to the energy density of the electric field of the electron ($r \to \infty$) we find $e^2 = a^2E = Al^{-2}a^2$. From Eqs. (14) and (15) the following relation between dimensionless parameters is found:

$$\alpha = e^2/\hbar c = a^2 R_0 l^{-1} I_1^{-1}. \tag{16}$$

As all of the numbers a, R_0/l, I_1 follow in principle from (7), Eq. (16) is a test on the validity of the assumption (2).[6] The following argument may also contribute to justify a correspondence between solitons and particles. The straight strings described have a constant "tension," a property which they share with the strings that have been introduced[7] in the context of quark confinement. It might be speculated that a soliton with $N = 3$ twists corresponds to the proton, which has a three-quark structure. As the validity of the model presented depends on the values of dimensionless constants and mass ratios, numerical solutions of Eq. (7) are desirable.

ACKNOWLEDGMENT

The author would like to thank Dr. K. Weiss of this laboratory for critically reading the manuscript.

[1] U. Enz. J. Math. Phys. **18**, 347 (1977).

[2] A further integer L has been introduced in Ref. 1, here we assume $L = 2$.

[3] S. S. Schweber, *Relativistic Quantum Field Theory* (Harper and Row, New York, 1962), p. 526.

[4] H. Hönal, Erg. Exact. Naturwiss. **26**, 291 (1952).

[5] L. de Broglie, *Non-Linear Wave Mechanics* (Elsevier, New York, 1960).

[6] In the possible alternative description based on closed strings of type (i), a fourth constant $K \neq 0$ is available, and the energy density of the electrical field can be separately adjusted. The present choice, however, is conceptically simpler and should be studied first. The exponent s' (here $s = 2$) is still open for adjustment.

[7] Y. Nambu, Lectures at the Copenhagen Summer Symposium, 1970 (unpublished).

Correlations in an N-mode laser[a)]

F. T. Hioe

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 1 November 1977)

Analytic expressions for the correlations of the light emitted by an N-mode laser are given. We study in particular a special case of strongest couplings and show, for example, that at steady state, the second order cross intensity correlation tends toward a negative value, $-1/(N-1)$, for large pump parameters. We also study the time dependent solutions, and express the second order amplitude and intensity correlations of the emitted light in terms of the eigenvalues and eigenfunctions of a Schrödinger-type differential equation which reduces to that studied by Risken for $N=1$ and to that studied by M-Tehrani and Mandel for $N=2$. Analytic expressions which approximate the eigenvalues of this differential equation for a general N are given.

1. INTRODUCTION

The theory of laser has been extensively studied for many years.[1] In the original theory of the laser put forward by Lamb,[2] he solved a set of coupled Maxwell equations for the field and Bloch equations for the atoms, and by demanding that the field be self-consistent, he obtained a set of coupled nonlinear equations in the field amplitudes. A very important generalization has been the inclusion to this set of equations of the random noises associated with the spontaneous atomic emission, for mathematically, this makes it possible to replace the deterministic coupled *nonlinear* equations in field amplitudes, $E_i(t)$, $i=1,2,\ldots,N$, by the probabilistic multidimensional Fokker–Planck equation *linear* in $p(E_1,E_2,\ldots,E_N;t)$, the probability of finding the field to be characterized by $E_1,E_2,\ldots,E_N$ at time t. Thus the inclusion of noises serves the doubly useful purposes of simplifying the problem mathematically as well as accounting for the naturally occurring effect physically. Moreover, all the physically interesting quantities such as the correlations $\langle E_i(t)E_j(t+\tau)\rangle$ can be obtained if the joint probability of finding the field to be characterized by $E_i(t)$ at time t and by $E_j(t+\tau)$ at time $t+\tau$ is known.

A number of papers based on the Fokker–Planck approach have been published. Of particular interest are perhaps those by Risken and Vollmer,[3] Hempstead and Lax,[4] and Wang and Lamb[5] for the one-mode case ($N=1$), Grossman and Richter,[6] and M-Tehrani and Mandel[7] for the two-mode case ($N=2$) and many others.[8]

In this paper, we present several analytic results for the general N-mode case, the general stationary solution among them. We shall be concerned in particular with what we shall refer to as the spherical case which is a special case of very strong couplings. We show, for example, that at steady state, the second-order cross intensity correlation tends toward the value $-1/(N-1)$ for large pump parameters. For $N=2$ this reduces to the result obtained by M-Tehrani and Mandel.[7] For the time-dependent solutions, we have expressed the second-order correlations of the emitted light in terms of the eigenvalues and eigenfunctions of a Schrödinger-type differential equation which reduces to that studied by Risken[3] for $N=1$ and to that studied by M-Tehrani and Mandel[7] for $N=2$. The compact expressions which we derived, as will be seen, are a result of our use of the hyperspherical coordinates which have been well studied but not very frequently used before. In the last section we present analytic representations of the eigenvalues and eigenfunctions which characterize the time-dependent correlation functions.

2. EQUATIONS OF MOTION AND THE STATIONARY STATE SOLUTION

The familiar equations of motion for the complex amplitudes $E_1(t),E_2(t),\ldots,E_N(t)$ of the N modes of a laser which were given by Lamb[9] and others are

$$\frac{dE_n}{dt}=a_nE_n-\sum_{\mu,\rho,\sigma=1}^{N}\sum E_\mu E_\rho E_\sigma \operatorname{Im}\{\theta_{n\mu\rho\sigma}\exp(i\Psi_{n\mu\rho\sigma})\},\quad n=1,2,\ldots,N, \tag{2.1}$$

where the a's are referred to as the pump parameters, and the θ's the coupling parameters, and the Ψ's are of the form

$$\Psi_{n\mu\rho\sigma}=(\nu_n-\nu_\mu+\nu_\rho-\nu_\sigma)t+\phi_n-\phi_\mu+\phi_\rho-\phi_\sigma. \tag{2.2}$$

We consider the so-called free running approximations[10] and we further supplement these equations by introducing random Langevin forces corresponding to spontaneous emission fluctuations. Equations (2.1) now take the form

$$\frac{dE_n}{dt}=E_n\left(a_n-\sum_{m=1}^{N}\theta_{nm}|E_m|^2\right)+\xi_n(t),\quad n=1,2,\ldots,N, \tag{2.3}$$

where the Langevin noise terms $\xi_1(t),\xi_2(t),\ldots,\xi_N(t)$ are assumed to be random and δ-correlated functions with zero mean values and

$$\langle\xi_i^*(t)\xi_j(t')\rangle=2\delta_{ij}\delta(t-t'). \tag{2.4}$$

It is known[11] that to a system of equations with fluctuating noises

$$\frac{dx_l}{dt}=f_l(\mathbf{x})+\sum_{m=1}^{d}g_{lm}(\mathbf{x})\xi_m(t),\quad l=1,2,\ldots,d, \tag{2.5}$$

where ξ's are independent, Gaussian, delta-correlated random functions with zero mean values and unit intensities, there corresponds a (multidimensional) Fokker–Planck equation for the joint probability density $p(x_1,x_2,\ldots,x_d;t)$ given by

a)Research partially supported by the Army Research Grant No. DAAG 29-77-G-0060.

 0022-2488/78/1906-1307$1.00

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(A_i p) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(B_{ij} p), \tag{2.6}$$

where

$$A_l(\mathbf{x}) = f_l(\mathbf{x}) + \sum_{m,j=1}^{d} \frac{\partial g_{lj}(\mathbf{x})}{\partial x_m} g_{mj}(\mathbf{x}), \tag{2.7}$$

$$B_{lm}(\mathbf{x}) = 2\sum_{j=1}^{d} g_{lj}(\mathbf{x}) g_{mj}(\mathbf{x}). \tag{2.8}$$

Thus, if we express the complex field amplitudes $E_1, E_2, \ldots, E_N$ in terms of real and imaginary parts

$$E_n = x_n + iy_n, \tag{2.9}$$

the $2N$-dimensional vector $\mathbf{x} \equiv (x_1, y_1, \ldots, x_N, y_N)$ represents the state of the laser, and the Fokker–Planck equation for the probability density $p(E_1, E_2, \ldots, E_N; t)$ or $p(\mathbf{x}, t)$ is given by

$$\frac{\partial p}{\partial t} = \sum_{i=1}^{N}\left[-\frac{\partial}{\partial x_i}(A_i^{(x)} p) - \frac{\partial}{\partial y_i}(A_i^{(y)} p) + \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2}\right) p\right], \tag{2.10}$$

where

$$A_i^{(\eta)} = \left[a_i - \sum_{m=1}^{N} \theta_{im}(x_m^2 + y_m^2)\right]\eta_i. \tag{2.11}$$

The introduction of the Langevin noise terms has thus conveniently led us to a linear equation in p from the more difficult nonlinear coupled equations in E's. Equation (2.10) was the starting point of the treatments of lasers by many authors.

The probability density $p_s(\mathbf{x})$ of the stationary state which the system generally reaches after a sufficiently long time is the time-independent solution of Eq. (2.10), i.e., p_s satisfies the equation

$$\sum_{i=1}^{N}\sum_{\eta=x,y} \frac{\partial}{\partial \eta_i}\left[(A_i^{(\eta)} p_s) - \frac{\partial}{\partial \eta_i} p_s\right] = 0 \tag{2.12}$$

which may be viewed as stating that the divergence of the probability current $A_i^{(\eta)} p_s - \partial p_s/\partial \eta_i$ vanishes. The probability current itself does not have to vanish in general. It has been shown[12] however, that the probability current itself vanishes, i.e.,

$$A_i^{(\eta)} p_s - \frac{\partial p_s}{\partial \eta_i} = 0 \tag{2.13}$$

in the special case called the potential case in which the $A_i^{(\eta)}(\mathbf{x})$'s are derivable from a potential function $U(\mathbf{x})$ such that

$$A_i^{(\eta)}(\mathbf{x}) = -\frac{\partial U(\mathbf{x})}{\partial \eta_i}, \tag{2.14}$$

and moreover, the $A_i^{(\eta)}(\mathbf{x})$'s satisfy

$$\frac{\partial A_i^{(\eta)}}{\partial \xi_j} = \frac{\partial A_j^{(\xi)}}{\partial \eta_i}, \quad \xi, \eta = x, y. \tag{2.15}$$

It can be verified that the potential function corresponding to Eq. (2.10) satisfying (2.15) is

$$U(\mathbf{x}) = \sum_{i=1}^{N}\left[-\tfrac{1}{2}a_i(x_i^2 + y_i^2) + \tfrac{1}{4}\theta_{ii}(x_i^2 + y_i^2)^2\right] + \tfrac{1}{4}\sum_{\substack{i,j=1 \\ i\neq j}}^{N}\sum \theta_{ij}(x_i^2 + y_i^2)(x_j^2 + y_j^2). \tag{2.16}$$

Thus the steady state probability density $p_s(\mathbf{x})$ is given by

$$p_s(\mathbf{x}) = Q^{-1}\exp[-U(\mathbf{x})], \tag{2.17}$$

where Q is the normalization constant given by

$$Q = \int_{-\infty}^{\infty}\cdots\int \exp[-U(\mathbf{x})]\, dx_1 dy_1 \cdots dx_N dy_N. \tag{2.18}$$

The expression for $U(\mathbf{x})$ for an N-mode laser, Eq. (2.16), has not appeared explicitly previously even though it is a rather straightforward generalization of those for the one-mode case given by Risken,[3] Hempstead and Lax,[4] and the two-mode case given by M-Tehrani and Mandel.[7]

We now consider the special case which we shall refer to as the spherical case in which

$$a_i = a, \quad \theta_{ij} = 1 \text{ for all } i,j = 1, 2, \ldots, N, \tag{2.19}$$

and we shall obtain moments of the light intensities for this case. We shall not use the polar coordinates

$$x_i = r_i\cos\phi_i, \quad y_i = r_i\sin\phi_i, \quad i = 1, 2, \ldots, N$$

as it was normally done, but we shall find it more convenient to use the hyperspherical coordinates[13]

$$\begin{aligned}
x_1 &= r\cos\theta_1,\\
y_1 &= r\sin\theta_1\cos\theta_2,\\
x_2 &= r\sin\theta_1\sin\theta_2\cos\theta_3,\\
y_2 &= r\sin\theta_1\sin\theta_2\sin\theta_3\cos\theta_4,\\
&\cdots\\
x_{N-1} &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{2N-4}\cos\theta_{2N-3},\\
y_{N-1} &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{2N-3}\cos\theta_{2N-2},\\
x_N &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{2N-2}\cos\phi,\\
y_N &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{2N-2}\sin\phi,
\end{aligned} \tag{2.20}$$

with $r \geq 0$, $0 \leq \theta_j \leq \pi$ $(j = 1, 2, \ldots, 2N-2)$, $0 \leq \phi \leq 2\pi$, except for the $N=1$ case for which we use, naturally, $x = r\cos\phi$, $y = r\sin\phi$, with $0 \leq \phi \leq 2\pi$. The volume element $dx_1 dy_1 \cdots dx_N dy_N$ in these hyperspherical coordinates is

$$r^{2N-1}(\sin\theta_1)^{2N-2}(\sin\theta_2)^{2N-3}\cdots(\sin\theta_{2N-2})\, dr\, d\theta_1\, d\theta_2 \cdots d\theta_{2N-2}\, d\phi. \tag{2.21}$$

It can be verified that

$$r^2 = \sum_{i=1}^{N}(x_i^2 + y_i^2). \tag{2.22}$$

Thus in the spherical case given by (2.19), the steady state probability density, (2.17), can be written as

$$p_s(r) = Q^{-1}\exp(\tfrac{1}{2}ar^2 - \tfrac{1}{4}r^4), \tag{2.23}$$

where

$$Q=\int_0^\infty r^{2N-1}\exp(\tfrac{1}{2}ar^2-\tfrac{1}{4}r^4)\,dr \times\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2N-2}(\sin\theta_2)^{2N-3}\cdots(\sin\theta_{2N-2}) \times d\theta_1\,d\theta_2\cdots d\theta_{2N-2}\,d\phi. \tag{2.24}$$

The integral over θ's and ϕ (the surface area) is $2\pi^N/\Gamma(N)$, where $\Gamma(x)$ is the Gamma function; and the integral over r can be expressed in terms of the confluent hypergeometric function,[14] ${}_1F_1(a,c;x)$ or $\Phi(a,c;x)$ in the Humbert's notation. We obtain

$$Q=\pi^{N+1/2}\left[\frac{1}{\Gamma(\frac{1}{2}(N+1))}\,\Phi\left(\frac{1}{2}N,\frac{1}{2};\frac{a^2}{4}\right)+\frac{a}{\Gamma(\frac{1}{2}N)} \times\Phi\left(\frac{1}{2}(N+1),\frac{3}{2};\frac{a^2}{4}\right)\right]. \tag{2.25}$$

The confluent hypergeometric function is a well studied function, and for convenience, some of its useful properties are listed in Appendix A. Using relations (A1), (A2), and (A3) of Appendix A, for example, we can express Q, for $N=1$ and 2 cases, in terms of the error function as given by Risken[3] and M-Tehrani and Mandel[7]:

$$N=1,\quad Q=\pi^{3/2}\exp(a^2/4)(1+\operatorname{erf}(a/2^{1/2})), \tag{2.26}$$

$$N=2,\quad Q=\pi^{5/2}[2/\pi^{1/2}+a\exp(a^2/4)(1+\operatorname{erf}(a/2))]. \tag{2.27}$$

Writing

$$I_i=x_i^2+y_i^2,\quad i=1,2,\ldots,N, \tag{2.28}$$

the moments of the light intensities at steady state

$$\langle I_{i_1}^{n_1}I_{i_2}^{n_2}\cdots\rangle=\int_{-\infty}^{\infty}\int(x_{i_1}^2+y_{i_1}^2)^{n_1}(x_{i_1}^2+y_{i_2}^2)^{n_2} \times\cdots p_s(\mathbf{x})\,dx_1\,dy_1\cdots dx_N\,dy_N \tag{2.29}$$

can be readily evaluated in the spherical case using the hyperspherical coordinate (2.20). Thus for the mean light intensity $\langle I_i\rangle$, since it is independent of i, we may choose to evaluate $\langle I_N\rangle$, and find

$$\langle I_N\rangle=Q^{-1}\int_0^\infty r^{2(N+1)-1}\exp(\tfrac{1}{2}ar^2-\tfrac{1}{4}r^4)\,dr\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2(N+1)-2}\cdots(\sin\theta_{2N-2})^3\,d\theta_1\cdots d\theta_{2N-2}\,d\phi$$
$$=\frac{\Phi(\frac{1}{2}(N+1),\frac{1}{2};a^2/4)/\Gamma(\frac{1}{2}(N+2))+a\Phi(\frac{1}{2}(N+2),\frac{3}{2};a^2/4)/\Gamma(\frac{1}{2}(N+1))}{\Phi(\frac{1}{2}N,\frac{1}{2};a^2/4)/\Gamma(\frac{1}{2}(N+1))+a\Phi(\frac{1}{2}(N+1),\frac{3}{2};a^2/4)/\Gamma(\frac{1}{2}N)}. \tag{2.30}$$

For large positive values of a, we obtain

$$\langle I_N\rangle\sim a/N \tag{2.31}$$

by applying relation (A4) in Appendix A to (2.30). It should perhaps be mentioned that although the third-order laser theory of Lamb holds in the neighborhood of the threshold, it covers a large range of pump parameters a of up to several hundred.

Similarly, we find

$$\langle I_N^2\rangle=Q^{-1}\int_0^\infty r^{2(N+2)-1}\exp(\tfrac{1}{2}ar^2-\tfrac{1}{4}r^4)\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2(N+2)-2}\cdots(\sin\theta_{2N-2})^5\,d\theta_1\cdots d\theta_{2N-2}\,d\phi$$
$$=2\times\frac{\Phi(\frac{1}{2}(N+2),\frac{1}{2};a^2/4)/\Gamma(\frac{1}{2}(N+3))+a\Phi(\frac{1}{2}(N+3),\frac{3}{2};a^2/4)/\Gamma(\frac{1}{2}(N+2))}{\Phi(\frac{1}{2}N,\frac{1}{2};a^2/4)/\Gamma(\frac{1}{2}(N+1))+a\Phi(\frac{1}{2}(N+1),\frac{3}{2};a^2/4)/\Gamma(\frac{1}{2}N)}\sim\frac{2a^2}{N(N+1)}\quad\text{as } a\to+\infty. \tag{2.32}$$

Also,

$$\langle I_{N-1}I_N\rangle=Q^{-1}\int_0^\infty r^{2(N+2)-1}\exp(\tfrac{1}{2}ar^2-\tfrac{1}{4}r^4)\,dr\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2(N+2)-2}\cdots(\sin\theta_{2N-4})^7$$
$$\times\sin^4\theta_{2N-3}\sin^3\theta_{2N-2}(\cos^2\theta_{2N-3}+\sin^2\theta_{2N-3}\cos^2\theta_{2N-2})\,d\theta_1\cdots d\theta_{2N-2}\,d\phi$$
$$=\tfrac{1}{2}\langle I_N^2\rangle\sim a^2/N(N+1)\quad\text{as } a\to+\infty. \tag{2.33}$$

Thus, denoting

$$\Delta I_i=I_i-\langle I_i\rangle, \tag{2.34}$$

we find, using (2.31), (2.32), and (2.33), that, for large positive values of a,

$$\langle(\Delta I_i)^2\rangle/\langle I_i\rangle^2\sim(N-1)/(N+1), \tag{2.35}$$

$$\langle(\Delta I_i)(\Delta I_j)\rangle/\langle I_i\rangle\langle I_j\rangle\sim-1/(N+1),\quad i\neq j, \tag{2.36}$$

$$\langle(\Delta I_i)(\Delta I_j)\rangle/[\langle(\Delta I_i)^2\rangle\langle(\Delta I_j)^2\rangle]^{1/2}\sim-1/(N-1),\quad i\neq j. \tag{2.37}$$

It is seen that for $N=2$, we recover the result of M-Tehrani and Mandel.[7]

The results above for the spherical case which represents a special case of strongest coupling, may be compared with the N independent-mode case specified by

$$\theta_{ij}=\delta_{ij} \tag{2.38}$$

in (2.3) and (2.10) in which we have

$$\langle I_i\rangle\sim a_i, \tag{2.39}$$

$$\langle I_i^2\rangle\sim a_i^2, \tag{2.40}$$

while $\langle(\Delta I_i)^2\rangle$ as well as all the cross correlation terms equal zero.

3. TIME-DEPENDENT SOLUTION

In this section, we consider the more general problem of finding the time-dependent solution for the probability density, $p(\mathbf{x}, t)$, of the Fokker–Planck equation (2.10). A substitution of

$$p(\mathbf{x}, t) = \exp[\psi(\mathbf{x})]g(\mathbf{x})T(t) \tag{3.1}$$

into (2.10), with the choice

$$\psi(\mathbf{x}) = -\tfrac{1}{2}U(\mathbf{x}) \tag{3.2}$$

where $U(\mathbf{x})$ is given by (2.16), results in the following eigenvalue equation:

$$T'/T = -\mathcal{L}g/g = -\lambda, \tag{3.3}$$

where

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{N}\sum_{\eta=x,y}\left(-\frac{\partial^2}{\partial\eta_i^2} + \frac{1}{4}(A_i^{(\eta)})^2 + \frac{1}{2}\frac{\partial A_i^{(\eta)}}{\partial\eta_i}\right)\\ &= \sum_{i=1}^{N}\left\{-\frac{\partial^2}{\partial x_i^2} - \frac{\partial^2}{\partial y_i^2} + \tfrac{1}{4}(x_i^2+y_i^2)\left[a_i - \sum_{m=1}^{N}\theta_{im}(x_m^2+y_m^2)\right]^2\right.\\ &\quad \left. + a_i - \theta_{ii}(x_i^2+y_i^2) - \sum_{m=1}^{N}\theta_{im}(x_m^2+y_m^2)\right\}.\end{aligned} \tag{3.4}$$

The factor $\exp[\psi(\mathbf{x})]$ in (3.1) and the choice (3.2) for $\psi(\mathbf{x})$ is to eliminate the first order derivative terms $\partial/\partial x_i$ and $\partial/\partial y_i$ such that the operator $\mathcal{L}$ given by (3.4) is a Sturm–Liouville self-adjoint operator.[15] If $\{n\}$ denotes the set of labels or "quantum numbers" for all the possible eigenvalues, the general solution of the Fokker–Planck equation (2.10) can be written as

$$p(\mathbf{x}, t) = \exp[-\tfrac{1}{2}U(\mathbf{x})]\sum_{\{n\}} c_{\{n\}}g_{\{n\}}(\mathbf{x})\exp(-\lambda_{\{n\}}t), \tag{3.5}$$

where the coefficients $c_{\{n\}}$ are to be determined by the boundary conditions. The lowest eigenvalue must be zero if the probability density $p(\mathbf{x}, t)$ is to reach a steady state limit, which has been shown to exist in the previous section. It also follows that the normalized eigenfunction $g_{\{0\}}(\mathbf{x})$ in

$$\mathcal{L}g = \lambda g \tag{3.6}$$

corresponding to the zero eigenvalue ($\lambda = 0$) is given by

$$g_{\{0\}}(\mathbf{x}) = Q^{-1/2}\exp(-\tfrac{1}{2}U(\mathbf{x})]. \tag{3.7}$$

It will be noted from Eqs. (2.17) and (3.7) that $p_s(\mathbf{x}) = |g_{\{0\}}(\mathbf{x})|^2$. More generally, we shall normalize all eigenfunctions of Eq. (3.6) such that

$$\int_{-\infty}^{\infty}\cdots\int g^*_{\{n\}'}(\mathbf{x})g_{\{n\}}(\mathbf{x})\,dx_1\,dy_1\cdots dx_N\,dy_N = \delta_{\{n\}\{n\}'}, \tag{3.8}$$

and we assume that they obey the completeness relation

$$\sum_{\{n\}} g_{\{n\}}(\mathbf{x}')g_{\{n\}}(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}'). \tag{3.9}$$

The conditional probability density, or the Green's function, $G(\mathbf{x}, t; \mathbf{x}_0, t_0)$ for the amplitudes of the fields to be characterized by $\mathbf{x}$ at time t if the amplitudes of the fields were characterized by $\mathbf{x}_0$ at time t_0, can be expressed in terms of the eigenvalues $\lambda_{\{n\}}$ and eigenfunctions $g_{\{n\}}$ as

$$\begin{aligned}G(\mathbf{x}, t; \mathbf{x}_0, t_0) &= \sum_{\{n\}} g^*_{\{n\}}(\mathbf{x})g_{\{n\}}(\mathbf{x}_0)\\ &\quad\times\exp[-\tfrac{1}{2}U(\mathbf{x}) + \tfrac{1}{2}U(\mathbf{x}_0)]\exp[-\lambda_{\{n\}}(t-t_0)],\\ &\qquad t \geq 0,\end{aligned} \tag{3.10}$$

for this function satisfies the original Fokker–Planck equation (2.10), and it reduces to $\delta(\mathbf{x}-\mathbf{x}_0)$ when $t = t_0$ according to (3.9). The joint probability density $p_2(\mathbf{x}, t; \mathbf{x}', t')$ for the field at two different times can now be written down. We have in the stationary state which is independent of the origin of time,

$$\begin{aligned}&p_2(\mathbf{x}, t+\tau; \mathbf{x}', t)\\ &\quad = G(\mathbf{x}, \tau; \mathbf{x}', 0)p_s(\mathbf{x}')\\ &\quad = Q^{-1}\sum_{\{n\}} g^*_{\{n\}}(\mathbf{x})g_{\{n\}}(\mathbf{x}')\exp[-\tfrac{1}{2}U(\mathbf{x}) - \tfrac{1}{2}U(\mathbf{x}')]\exp(-\lambda_{\{n\}}\tau),\\ &\qquad\qquad \tau \geq 0\\ &\quad = g^*_{\{0\}}(\mathbf{x})g_{\{0\}}(\mathbf{x}')\sum_{\{n\}} g^*_{\{n\}}(\mathbf{x})g_{\{n\}}(\mathbf{x}')\exp(-\lambda_{\{n\}}\tau),\\ &\qquad\qquad \tau \geq 0.\end{aligned} \tag{3.11}$$

The second-order amplitude and intensity correlations defined by

$$\begin{aligned}&\langle E^*_{j'}(t)E_j(t+\tau)\rangle\\ &\quad = \int_{-\infty}^{\infty}\cdots\int (x'_{j'} - iy'_{j'})(x_j + iy_j)p_2(\mathbf{x}, t+\tau; \mathbf{x}', t)d^{2N}x\,d^{2N}x',\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}&\langle I_{j'}(t)I_j(t+\tau)\rangle\\ &\quad = \int_{-\infty}^{\infty}\cdots\int (x_{j'}'^2 + y_{j'}'^2)(x_j^2 + y_j^2)p_2(\mathbf{x}, t+\tau; \mathbf{x}', t)d^{2N}x\,d^{2N}x',\end{aligned} \tag{3.13}$$

where $d^{2N}x \equiv dx_1\,dy_1\cdots dx_N\,dy_N$, can be expressed in terms of the eigenvalues and eigenfunctions of the operator $\mathcal{L}$ in (3.4).

We shall now consider the spherical case as specified by (2.19). Noting the definition of r given by Eq. (2.22), the operator $\mathcal{L}$ in Eq. (3.4) becomes

$$\mathcal{L} = \sum_{i=1}^{N}\left(-\frac{\partial^2}{\partial x_i^2} - \frac{\partial^2}{\partial y_i^2}\right) + F(r), \tag{3.14}$$

where

$$F(r) = Na - (N+1)r^2 + \tfrac{1}{4}r^2(a-r^2)^2. \tag{3.15}$$

If we use the hyperspherical coordinates (2.20) and write the eigenfunctions, $g(\mathbf{x})$, of $\mathcal{L}$ in the form

$$g(\mathbf{x}) = R(r)Y(\theta_1, \theta_2, \ldots, \theta_{2N-2}, \phi) \tag{3.16}$$

the radial part, $R(r)$, of $g(\mathbf{x})$ satisfies the equation[13]

$$\frac{d^2R_{nl}}{dr^2} + \frac{2N-1}{r}\frac{dR_{nl}}{dr} + \left[\lambda_{nl} - F(r) - \frac{l(l+2N-2)}{r^2}\right]R_{nl} = 0, \tag{3.17}$$

while the Y's are known functions [independent of $F(r)$ as F does not depend on θ's or ϕ] which can be expressed in terms of the Gegenbauer polynomials.[13] The eigenvalues λ_{nl} and the radial eigenfunctions $R_{nl}(r)$ depend, of course, on N but the dependence is not explicitly shown for notational convenience. For every value of $l = 0, 1, 2, \cdots$ (which can be referred to as the orbital quantum number), there are $h(N, l)$ linearly independent Y's, where

$$h(N, l) = 2(N+l-1)\cdot\frac{(2N+l-3)!}{(2N-2)!\,l!}. \tag{3.18}$$

The $h(N, l)$ linearly independent Y's are given by[13]

$$Y_{\{m\}}(\theta_1, \theta_2, \ldots, \theta_{2N-2}, \phi)$$
$$= M(m_0, m_1, \ldots, m_{2N-2}) \exp(\pm i m_{2N-2}\phi) \prod_{k=0}^{2N-3} (\sin\theta_{k+1})^{m_{k+1}}$$
$$\times C^{m_{k+1}+N-k/2-1}_{m_k-m_{k+1}}(\cos\theta_{k+1}), \tag{3.19}$$

where $m_0, m_1, \ldots, m_{2N-2}$ are integers such that $l = m_0 \geq m_1 \geq \ldots \geq m_{2N-2} \geq 0$, and $C^{\nu}_n(x)$ is the Gegenbauer polynomial of degree n and order ν, and $M(m_0, m_1, \ldots, m_{2N-2})$ is some normalization constant which we shall use later [see Eq. (3.27)].

Returning to Eq. (3.17), the first order derivative term in the equation can be eliminated by the following substitution

$$R(r) = r^{-(2N-1)/2}\chi(r) \tag{3.20}$$

which results in the following equation:

$$\frac{d^2\chi_{nl}}{dr^2} + \left[\lambda_{nl} - F(r) - \frac{(l+N-\frac{3}{2})(l+N-\frac{1}{2})}{r^2}\right]\chi_{nl} = 0, \tag{3.21}$$

where $F(r)$ is given by (3.15). Equation (3.21) gives the relevant eigenvalue equation for any number of modes in the spherical case in a rather compact form. It can be verified that for $N=1$ and 2, it reduces to those equations studied by Risken,[3] and by M-Tehrani and Mandel.[7]

We have labeled the eigenvalues of Eq. (3.21) λ_{nl} since for every value of l, we have $n = 0, 1, 2, \ldots$ (n may be referred to as the principal quantum number). Each of these eigenvalues is generally degenerate with $h(N,l)$ given by (3.18) as the number of degeneracies, since there are $h(N,l)$ independent eigenfunctions corresponding to a given λ_{nl}. Thus in writing the joint probability density, Eq. (3.11), as

$$p_2(\mathbf{x}, t+\tau; \mathbf{x}', t)$$
$$= g^*_{00}(\mathbf{x}) g_{00}(\mathbf{x}') \sum_{n,l=0}^{\infty} g^*_{nl}(\mathbf{x}) g_{nl}(\mathbf{x}') \exp(-\lambda_{nl}\tau), \quad \tau \geq 0, \tag{3.22}$$

we should remember to take into account of the appropriate degeneracies. The "ground state" eigenfunction $g_{00}(\mathbf{x})$ corresponding to $\lambda_{00} = 0$ is, from Eq. (3.7), known for a general set of parameter a_i and θ_{ij}. For the spherical case in particular, the normalized ground state eigenfunction is

$$g_{00}(\mathbf{x}) = Q^{-1/2} \exp(\tfrac{1}{4}ar^2 - \tfrac{1}{8}r^4), \tag{3.23}$$

where Q is given by (2.25).

The exact analytic form of the general $g_{nl}(\mathbf{x})$ and λ_{nl} other than the ground state is not known, but we have found simple approximate analytic expressions for them and these will be presented in the following section. In the remainder of this section, we want to show that for the second-order amplitude and intensity correlations (3.12) and (3.13), not all eigenvalues and eigenfunctions contribute to them. The "selection rule" arises, as we shall see, when we integrate the angular parts of the eigenfunctions.

First, let us write out (3.12) and (3.13) more explicitly. For the spherical case, as pointed out in the previous section, because of the symmetry, these correlations do not depend on the j's. Therefore, without any loss of generality, for the case $j' = j$, we may choose $j = N$, and for the case $j' \neq j$, we may choose $j' = N-1$, $j = N$. We use the hyperspherical coordinates (2.20) and consider first the case $j' = j$, for which, following (3.11), (3.12), (3.16), and (3.20), we find

$$\langle E^*_j(t) E_j(t+\tau)\rangle = \sum_{n,l=0}^{\infty} u_{nl} \exp(-\lambda_{nl}\tau), \tag{3.24}$$

where

$$u_{nl} = \Big| \int_0^{\infty} r\chi_{00}(r)\chi_{nl}(r)\,dr \int_0^{\pi} \cdots \int_0^{\pi}\int_0^{2\pi} (\sin\theta_1)^{2N-1}$$
$$\times (\sin\theta_2)^{2N-2} \cdots (\sin\theta_{2N-2})^2 \exp(i\phi)$$
$$\times Y_0 Y_l(\theta_1, \theta_2, \ldots, \theta_{2N-2}, \phi)\, d\theta_1 d\theta_2 \cdots d\theta_{2N-2} d\phi \Big|^2. \tag{3.24a}$$

The appropriate degeneracies for every l value should always be remembered. The normalization of the eigenfunctions $g(\mathbf{x})$, Eq. (3.8), means that the transformed eigenfunctions $\chi_{nl}(r)$ given by (3.21) have been normalized such that

$$\int_0^{\infty} |\chi_{nl}(r)|^2\, dr = 1 \tag{3.25}$$

and that the Y_l's have been normalized such that

$$\int_0^{\pi} \cdots \int_0^{\pi}\int_0^{2\pi} (\sin\theta_1)^{2N-2} (\sin\theta_2)^{2N-3} \cdots \sin\theta_{2N-2}$$
$$\times |Y_l(\theta_1, \theta_2, \ldots, \theta_{2N-2}\phi)|^2\, d\theta_1 d\theta_2 \cdots d\theta_{2N-2} d\phi = 1. \tag{3.26}$$

To satisfy (3.26), M in (3.19) should be chosen to be (Ref. 13, p. 240)

$$M(m_0, m_1, \ldots, m_{2N-2}) = \left\{2\pi \prod_{k=1}^{2N-2} E_k(m_{k-1}, m_k)\right\}^{-1/2}, \tag{3.27}$$

where

$$E_k(l,m) = \frac{\pi 2^{k-2m-2N+2}\Gamma(l+m+2N-1-k)}{(l+N-\frac{1}{2}-\frac{1}{2}k)(l-m)!\,[\Gamma(m+N-\frac{1}{2}-\frac{1}{2}k)]^2}. \tag{3.28}$$

Let us examine the integral over the angular part in (3.24) first for the case $N=1$. The integral in this case is

$$\int_0^{2\pi} \exp[i(1 \pm l)\phi]\, d\phi \tag{3.29}$$

from which it is clear that only $l=1$, among all nonnegative values of l, has a nonzero contribution in (3.24).

Next we consider the $N=2$ case. The integral over the angular part in this case is

$$\int_0^{2\pi} \exp[i(1 \pm m_2)\phi]\, d\phi \int_0^{\pi} (\sin\theta_1)^{m_1+3} C^{m_1+3}_{l-m_1}(\cos\theta_1)\, d\theta_1$$
$$\times \int_0^{\pi} (\sin\theta_2)^{m_2+2} C^{m_2+1/2}_{m_1-m_2}(\cos\theta_2)\, d\theta_2. \tag{3.30}$$

We see that, to be nonzero, $m_2 = 1$. The integral over θ_2 then becomes

$$\int_{-1}^{1} (1-x^2) C^{3/2}_{m_1-1}(x)\, dx = \int_{-1}^{1} (1-x^2)^{3/2-1/2} C^{3/2}_{m_1-1}(x)\, dx \tag{3.31}$$

which, from the orthogonality property of the Gegenbauer polynomial [see Eq. (B1) in Appendix B], is zero unless $m_1 = 1$. The integral over θ_1 then becomes

$$\int_{-1}^{1} (1-x^2)^{3/2} C^2_{l-1}(x)\, dx = \int_{-1}^{1} (1-x^2)^{2-1/2} C^2_{l-1}(x)\, dx \tag{3.32}$$

from which it follows that only $l=1$ contributes in (3.24) for the case $N=2$.

For the more general value of $N(>2)$, other values of l will generally contribute. All that can be said is that the only contributing value of m_{2N-4}, m_{2N-3}, and m_{2N-2} in the set $l=m_0\geq m_1\geq\cdots\geq m_{2N-2}\geq 0$ is 1.

For $j'\neq j$, $\langle E_{j'}^*(t)E_j(t+\tau)\rangle=0$. This follows readily if we note that the integral over ϕ and ϕ' as we choose $j'=N-1$ and $j=N$, is given by

$$\int_0^{2\pi}\exp(\pm im_{2N-2}\phi')d\phi'\int_0^{2\pi}\exp[i(1\pm m_{2N-2})\phi]\,d\phi=0. \tag{3.33}$$

Next consider the intensity correlation for the case $j'=j$, which, following (3.11), (3.13), (3.16) and (3.20), can be expressed as

$$\langle I_j(t)I_j(t+\tau)\rangle=\sum_{n,\,l=0}^{\infty}v_{nl}\exp(-\lambda_{nl}\tau), \tag{3.34}$$

where

$$\begin{aligned}v_{nl}=\Big|&\int_0^\infty r^2\chi_{00}(r)\chi_{nl}(r)\,dr\\&\times\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2N}(\sin\theta_2)^{2N-1}\cdots(\sin_{2N-2})^3\\&\times Y_0Y_l(\theta_1,\theta_2,\ldots,\theta_{2N-2},\phi)\,d\theta_1d\theta_2\cdots d\theta_{2N-2}d\phi\Big|^2.\end{aligned} \tag{3.34a}$$

First consider the case $N=1$. The integral over the angular part is

$$\int_0^{2\pi}\exp(\pm il\phi)\,d\phi$$

which shows that only $l=0$ contributes in (3.34). For $N=2$, the integral over the angular part is

$$\int_0^{2\pi}\exp(\pm im_2\phi)\,d\phi\int_0^\pi(\sin\theta_1)^{m_1+4}C_{l-m_1}^{m_1+1}(\cos\theta_1)\,d\theta_1\times\int_0^\pi(\sin\theta_2)^{m_2+3}C_{m_1-m_2}^{m_2+1/2}(\cos\theta_2)\,d\theta_2.$$

To be nonzero, first $m_2=0$. The integral over θ_2 is then

$$\int_{-1}^1(1-x^2)C_{m_1}^{1/2}(x)\,dx=\int_{-1}^1(1-x^2)(1-x^2)^{1/2-1/2}C_{m_1}^{1/2}(x)\,dx, \tag{3.35}$$

which is zero unless $m_1=0$ or 2. When $m_1=0$, the integral over θ_1 is

$$\int_{-1}^1(1-x^2)^{3/2}C_l^1(x)\,dx=\int_{-1}^1(1-x^2)(1-x^2)^{1-1/2}C_l^1(x)\,dx, \tag{3.36}$$

which is zero unless $l=0$ or 2. When $m_1=2$, the integral over θ_1 is

$$\int_{-1}^1(1-x^2)^{5/2}C_{l-2}^3(x)\,dx=\int_{-1}^1(1-x^2)^{3-1/2}C_{l-2}^3(x)\,dx, \tag{3.37}$$

which is zero unless $l=2$. We thus conclude that for the case $N=2$, $l=0$, and 2 are the only contributing values of l for the second order intensity correlations (3.33).

For the more general value of $N(>2)$, other values of l will in general contribute. We can say, however, that in the set of m values $l=m_0\geq m_1\geq\cdots\geq m_{2N-2}\geq 0$, the only contributing values of m_{2N-4} and m_{2N-3} are 0 and 2 and of m_{2N-2} zero only.

For the case $j'\neq j$, $\langle I_{j'}(t)I_j(t+\tau)\rangle$ is given by

$$\langle I_{j'}(t)I_j(t+\tau)\rangle=\sum_{n,\,l=0}^{\infty}w_{nl}\exp(-\lambda_{nl}\tau), \tag{3.38}$$

where

$$\begin{aligned}w_{nl}=\Big|&\int_0^\infty r^2\chi_{00}(r)\chi_{nl}(r)\,dr\Big|^2\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2N}(\sin\theta_2)^{2N-1}\cdots(\sin\theta_{2N-2})Y_0Y_l(\theta_1,\ldots,\theta_{2N-2}\phi)\,d\theta_1\cdots d\theta_{2N-2}d\phi\\&\times\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi}(\sin\theta_1)^{2N}(\sin\theta_2)^{2N-1}\cdots(\sin\theta_{2N-4})^5(\cos^2\theta_{2N-3}+\sin^2\theta_{2N-3}\cos^2\theta_{2N-2})\sin^2\theta_{2N-3}\sin\theta_{2N-2}\\&\times Y_0Y_l(\theta_1,\ldots,\theta_{2N-2}\phi)\,d\theta_1\cdots d\theta_{2N-2}d\phi.\end{aligned} \tag{3.38a}$$

For $N=2$, it can be shown similarly that $l=0$ and 2 are the only contributing values of l for this correlation.

For easy reference, some of the useful properties of the Gegenbauer polynomial[16] $C_n^\nu(x)$ are listed in Appendix B. Using these properties together with the well known formula

$$\int_0^{\pi/2}\sin^{m-1}\theta\cos^{n-1}\theta\,d\theta=\frac{1}{2}\frac{\Gamma(m/2)\Gamma(n/2)}{\Gamma[(m+n)/2]}, \tag{3.39}$$

it can be verified that the exact expressions for the cases $N=1$ and 2 can be written more simply as follows:

$N=1$

$$\langle E^*(t)E(t+\tau)\rangle=\sum_{n=0}^{\infty}\exp(-\lambda_{n1}\tau)\Big|\int_0^\infty r\chi_{00}(r)\chi_{n1}(r)\,dr\Big|^2, \tag{3.40}$$

$$\langle I(t)I(t+\tau)\rangle=\sum_{n=0}^{\infty}\exp(-\lambda_{n0}\tau)\Big|\int_0^\infty r^2\chi_{00}(r)\chi_{n0}(r)\,dr\Big|^2, \tag{3.41}$$

$N=2$

$$\langle E_{j'}^*(t)E_j(t+\tau)\rangle=\begin{cases}\sum_{n=0}^{\infty}\exp(-\lambda_{n1}\tau)\Big|\int_0^\infty r\chi_{00}(r)\chi_{n1}(r)\,dr\Big|^2, & j'=j, \quad (3.42a)\\ 0, & j'\neq j, \quad (3.42b)\end{cases}$$

$$\langle I_{j'}(t)I_j(t+\tau)\rangle=\tfrac{1}{4}\sum_{n=0}^{\infty}\Big\{\exp(-\lambda_{n0}\tau)\Big|\int_0^\infty r^2\chi_{00}(r)\chi_{n0}(r)\,dr\Big|^2\pm\tfrac{1}{3}\exp(-\lambda_{n2}\tau)\Big|\int_0^\infty r^2\chi_{00}(r)\chi_{n2}(r)\,dr\Big|^2\Big\}, \tag{3.43}$$

where the + sign is for the case $j'=j$ and the − sign is for the case $j'\neq j$.

The result for $N=2$ agrees with that obtained by M-Tehrani and Mandel. Some numerical data and graphs for various quantities of interest for this case have been presented in their paper.

The ground state eigenvalue, λ_{00}, is, as already shown, exactly zero independent of the value of the

pump parameter a. It has been observed numerically, for the cases of $N=1$ and 2, that the values of λ_{0l} for $l\neq 0$ approach zero as the pump parameter a is increased and becomes positively large. This means that for large pump parameter, the intensity correlation for the case $N=2$ is no longer well represented by a single exponential as in the $N=1$ case and that the correlations persist for longer times. This also holds for $N>2$. An argument which indicates that $\lambda_{0l}\to 0$ as a becomes very large for any value of N will be given in the next section.

4. ANALYTIC APPROXIMATIONS OF THE EIGENVALUES λ_{nl}

In the previous section, we have expressed the second-order amplitdue and intensity correlations, (3.12) and (3.13), in terms of the eigenvalues and eigenfunctions of the operator $\mathcal{L}$ given in (3.4). For the spherical case specified by (2.19), the correlations (3.24), (3.34), and (3.38) are expressed in terms of the eigenvalues λ_{nl} and eigenfunctions $\chi_{nl}(r)$ of Eq. (3.21). The exact analytic solution of Eq. (3.21) is not known, and previous studies for the cases $N=1$ and 2 have been based on computer numerical calculations.

Differential equations similar to Eq. (3.21) have been recently studied by the author[17–19] in collaboration with Montroll and MacMillen in a different problem, and we have found that the WKB formula, which in theory provides good approximations to the eigenvalues when the quantum number is large, is good even when the quantum number is small. We shall see in this section that when properly scaled and expanded, the WKB formula provides a beautifully simple formula for all eigenvalues and all values of parameters. We shall also indicate how the eigenfunctions can be analytically represented.

The basis of our analytic approximation scheme is Titchmarsh's formula[20] which is a rigorous version of the WKB formula. Beginning with Eq. (3.21), Titchmarsh's theorem states that the eigenvalues λ_{nl}, in the large quantum number limit, are given by

$$(1/\pi)\int_0^{r_0}[\lambda_{nl}-F(r)]^{1/2}dr=n+\tfrac{1}{2}(l+N), \tag{4.1}$$

where r_0 is the smallest positive root of

$$\lambda_{nl}-F(r_0)=0. \tag{4.2}$$

With $F(r)$ given by (3.15), the integral on the left-hand side of (4.1) can be expressed in terms of complete elliptic integrals of the first, second, and third kinds.[18] However, this expression is very lengthy and too cumbersome to be used for practical purposes. As shown in Ref. 19 on the other hand, the integral on the left-hand side of (4.1) can be expanded in simple series either in integral powers of λ_{nl} for the case of small λ_{nl} or in integral powers of $\lambda_{nl}^{-1/3}$ for the case of large λ_{nl}. The appearance of the 1/3 power is interesting and is due to a remarkably useful scale transformation. The reader is referred to Ref. 19 for detail. The large λ_{nl} expansion turns out to cover most cases for our problem here.

Denoting

$$\Lambda_{nl}\equiv\lambda_{nl}-Na, \tag{4.3}$$

we find, following the expansion procedure described in Ref. 19, that Eq. (4.1) becomes

$$\frac{\Lambda_{nl}^{2/3}}{6\cdot 2^{2/3}\pi^{1/2}}\sum_{j=0}^{\infty}B_j=n+\tfrac{1}{2}(l+N), \tag{4.4}$$

where

$$B_j=\frac{(-1)^j}{j!}\sum_{p=0}^{j}\binom{j}{p}\beta^{j-p}\gamma^p\frac{\Gamma((2j+2p+1)/6)}{\Gamma((5+j+p)/3-j)}, \tag{4.5}$$

$$\beta=2^{2/3}(a^2/4-N-1)\Lambda_{nl}^{-2/3},\quad \gamma=-2^{1/3}a\Lambda_{nl}^{-1/3}. \tag{4.6}$$

$|\Lambda_{nl}|$ must be greater than the larger of

$$2|a|^3/27 \quad\text{and}\quad 2|(a^2/4-N-1)^{\frac{2}{3}}|^{3/2} \tag{4.7}$$

for the series $\sum B_j$ to converge. By successively truncating the series on the left-hand side of (4.4), successive approximations of λ_{nl} for any values of n, l, and N can be obtained. Formula (4.4) can be written out more explicitly in the form

$$\Lambda_{nl}^{2/3}(b_0+b_1\Lambda_{nl}^{-1/3}+b_2\Lambda_{nl}^{-2/3}+b_3\Lambda_{nl}^{-1}+b_4\Lambda_{nl}^{-4/3}+\cdots)$$
$$=n+\tfrac{1}{2}(l+N), \tag{4.8}$$

where we found

$$b_0=0.36525,$$
$$b_1=0.094341a,$$
$$b_2=(N+1)/6,$$
$$b_3=0.081167(N+1)a-0.0045093a^3,$$
$$b_4=0.031447(N+1)^2+0.010482(N+1)a^2$$
$$-0.00058235a^4. \tag{4.9}$$

By inverting this expression, λ_{nl} can be expressed explicitly as follows:

Denoting

$$\mu\equiv n+\tfrac{1}{2}(l+N) \tag{4.10}$$

we have

$$\lambda_{nl}=Na+4.53018\mu^{3/2}$$
$$\times[1+A_1\mu^{-1/2}+A_2\mu^{-1}+A_3\mu^{-3/2}+A_4\mu^{-2}+\cdots]^3, \tag{4.11}$$

where

$$A_1=-0.078051a,$$
$$A_2=-(N+1)/12+0.0030459a^2,$$
$$A_3=-0.024527(N+1)a+0.0013626a^3,$$
$$A_4=-0.0092152(N+1)^2-0.0035748(N+1)a^2$$
$$+0.00020807a^4. \tag{4.12}$$

Some samples of λ_{nl} for $l=0$, $a=0$, obtained from this formula (with terms up to A_4 only) are compared in Table I with the numerical results obtained by Risken ($N=1$) and by the author ($N=2$). The accuracy is seen to be quite remarkable. Formula (4.11) has been found to be also reasonably good (always increasingly better, of course, as n,l increase) for other values of parameters; we must remember, however, the range of its validity as specified by (4.7).

TABLE I. Comparison of some λ_{nl} for $l=a=0$ obtained numerically and from Eq. (4.11).

	$N=1$		$N=2$	
n	Numerical	Eq. (4.11)	Numerical	Eq. (4.11)
1	5.6266	5.5279	7.644	7.727
2	14.3628	14.2849	17.38	17.457
3	25.4522	25.3820	29.17	29.285
4	38.4622	38.3970	42.73	42.912
5	53.1453	53.0798	57.85	58.133

The accuracy of our analytic expression (4.11) for λ_{nl} can be further improved if we make use of the exact numerical data obtained by other methods. One procedure for such improvement was given in Refs. 17 and 18.

Our formula (4.11) can also be readily adapted for the N independent one-mode case specified by (2.38). Denoting the A_i in (4.12) more specifically by $A_i(N,a)$, and the μ in (4.10) by $\mu(n,l,N)$, then the eigenvalues for the N independent one-mode case can be approximated by

$$\lambda_{n_1 l_1 \cdots n_N l_N}(a_1,\ldots,a_N) = \sum_{p=1}^{N}\left\{a_p + 4.53018\,\mu(n_p,l_p,1)^{3/2} \times\left[1+\sum_{i=1}^{\infty} A_i(1,a_p)\,\mu(n_p,l_p,1)^{-i/2}\right]^3\right\}. \tag{4.13}$$

The spherical case and the N independent one-mode case are two extreme cases of the general situation in which the sets of a_i and θ_{ij} lie between these two extreme cases. The study of the more general cases will involve the study of multidimensional differential equations. This will be taken up in another paper in which we shall approach the problem with techniques similar to those used in Ref. 19.

The eigenfunctions $\chi_{nl}(r)$ of Eq. (3.21) can also be analytically represented quite accurately if we combine the WKB type analysis with the exact numerical data. For example, by observing the numerical data, we have found that the normalized eigenfunctions $\chi_{nl}(r)$ for $n\neq 0$ [the ground state eigenfunction

$$\chi_{00}(r) = [2\pi^N/\Gamma(N)]^{1/2} Q^{-1/2} r^{(2N-1)/2}\exp(\tfrac{1}{4}ar^2 - \tfrac{1}{8}r^4)$$

exactly from (3.20), (3.23), (2.24), and (3.25) for any N] can be quite accurately represented by the following simple form:

$$\chi_{nl}(r) = \begin{cases} c_{nl}\cos(B\lambda_{nl}^{1/2} r - \tfrac{1}{2}\pi[l+N-\tfrac{1}{2}]), & r_{\min}^{(nl)} \le r \le r_{\max}^{(nl)}, \\ 0, & \text{otherwise}, \end{cases} \tag{4.14}$$

where

$$r_{\min}^{(nl)} = (\pi/2B\lambda_{nl}^{1/2})\,(l+N-\tfrac{3}{2}), \tag{4.15}$$

$$r_{\max}^{(nl)} = (\pi/2B\lambda_{nl}^{1/2})(l+N+2n+\tfrac{1}{2}), \tag{4.16}$$

$$c_{nl} = [2B\lambda_{nl}^{1/2}/(n+1)\pi]^{1/2}, \tag{4.17}$$

λ_{nl} is the eigenvalue [approximately given by Eq. (4.11)] and B is found numerically to be ≈ 1. The actual eigenfunctions $\chi_{nl}(r)$ rise slowly from zero at $r=0$ and decrease exponentially to zero at the tail (after cutting the r axis n times), rather than becoming zero abruptly outside the range $r_{\min}^{(nl)} \le r \le r_{\max}^{(nl)}$ as given by (4.14). The error introduced for any practical purpose is, however, very small. Notice that n is the "principal quantum number" which gives the number zeros of the eigenfunction $\chi_{nl}(r)$ between $r_{\min}^{(nl)}$ and $r_{\max}^{(nl)}$. Further improvements of this approximation can be made. Details of this investigation of the analytic representations of the eigenfunctions will again be left to the next paper.

Formula (4.11), by the restriction (4.7), does not cover the case when the value of the eigenvalue is very small. As pointed out in the previous section, aside from $\lambda_{00}=0$, λ_{0l} for $l>0$, has been found to approach zero for large positive values of pump parameter a. This can be seen as follows:

For very large a, the minimum of the potential

$$V(r) = \frac{(l+N-\frac{3}{2})(l+N-\frac{1}{2})}{r^2} + Na - (N+1)r^2 + \tfrac{1}{4}r^2(a-r^2)^2 \tag{4.18}$$

in Eq. (3.21) can be verified to occur at $r^2=a$. The eigenfunction $\chi_{0l}(r)$, since $n=0$, is a bell-shaped function having only one peak which should occur at approximately $r^2=a$. For $l=0$, we know that $\chi_{00}(r)$ is exactly given by

$$\chi_{00}(r) = c r^{(2N-1)/2}\exp(\tfrac{1}{4}ar^2 - \tfrac{1}{8}r^4). \tag{4.19}$$

For $l\neq 0$ it can be verified that

$$\chi_{0l}(r) = c r^{[(2N-1)/2]+l}\exp(\tfrac{1}{4}ar^2 - \tfrac{1}{8}r^4) \tag{4.20}$$

satisfies exactly the differential equation (3.21) with $\lambda_{0l}=0$ at $r^2=a$, although only approximately elsewhere. Since the most important feature of the eigenfunction is around $r^2=a$, it is understandable, as the numerical result shows, that, for large a, $\lambda_{0l}\approx 0$ with the corresponding eigenfunction given approximately by (4.20).

5. SUMMARY

We have presented some analytic results for an N-mode laser, among them the steady state probability density (2.17) for a rather general case, and in the special case which we called the spherical case, the steady state correlations (2.35)–(2.37), and the time dependent correlations (3.24), (3.34), and (3.38). The eigenvalue equation, the solution of which the time-dependent correlations depend on, is expressed in a compact form by Eq. (3.21). The $N=1$ and 2 results reduce to those obtained previously by other authors. We have also presented a simple and useful analytic representation of the eigenvalues (4.11) which has been found to be reasonably accurate for a wide range of parameters.

ACKNOWLEDGMENT

The author is very grateful to Mohammad M-Tehrani and Leonard Mandel for stimulating his interest in this research.

Note added in manuscript: The author is pleased to learn that the recent experimental results obtained by Mandel and M-Tehrani for a two-mode laser (to be

published in Opt. Lett.) agreed well with the theoretical calculations given by us.

APPENDIX A

In this appendix, we list some of the useful properties of the confluent hypergeometric function,[14] ${}_1F_1(a,c;x) \equiv \Phi(a,c;x)$:

$$\frac{2x}{\pi^{1/2}}\Phi(\tfrac{1}{2},\tfrac{3}{2};-x^2) = \operatorname{erf}(x) \equiv \frac{2}{\pi^{1/2}}\int_0^x e^{-t^2}\,dt = 1 - \operatorname{erfc}(x), \tag{A1}$$

$$\Phi(a,c;x) = e^x\Phi(c-a,c,-x), \tag{A2}$$

$$\frac{d}{dx}\Phi(a,c;x) = \frac{a}{c}\Phi(a+1,c+1;x)$$

$$= \frac{a}{x}\{\Phi(a+1,c;x) - \Phi(a,c;x)\}, \tag{A3}$$

$$\Phi(a,c;x) = \frac{\Gamma(c)}{\Gamma(a)}e^x x^{a-c}[1 + O(|x|^{-1})] \quad \text{as } \mathrm{Re}x \to \infty. \tag{A4}$$

APPENDIX B

In this appendix, we list some of the useful properties of the Gegenbauer polynomial[16] $C_n^\alpha(x)$:

$$\int_{-1}^{1}(1-x^2)^{\alpha-1/2}C_m^\alpha(x)C_n^\alpha(x)\,dx = \delta_{mn}\frac{\pi 2^{1-2\alpha}\Gamma(n+2\alpha)}{n!\,(n+\alpha)[\Gamma(\alpha)]^2}, \quad \alpha \neq 0, \quad \alpha > -\tfrac{1}{2}, \tag{B1}$$

$$C_n^\alpha(x) = \frac{1}{\Gamma(\alpha)}\sum_{m=0}^{[n/2]}(-1)^m\frac{\Gamma(\alpha+n-m)}{m!\,(n-2m)!}(2x)^{n-2m}, \quad \alpha > -\tfrac{1}{2}, \quad \alpha \neq 0, \tag{B2}$$

$$C_n^{(0)}(x) = \sum_{m=0}^{[n/2]}(-1)^m\frac{(n-m-1)!}{m!\,(n-2m)!}(2x)^{n-2m}. \tag{B3}$$

[1] See M. Sargent III, M.O. Scully, and W.E. Lamb, Jr. (hereafter called SSL), *Laser Physics* (Addison-Wesley, Reading, Mass., 1974); and H. Haken, "Laser Theory," in *Handbuch der Physik*, edited by S. Flügge (Springer, New York, 1970), Vol. XXV/2c, and the extensive references therein.
[2] W.E. Lamb, Jr., Phys. Rev. A **134**, 1429 (1964).
[3] H. Risken, Z. Phys. **191**, 186 (1965); H. Risken and H.D. Vollmer, Z. Phys. **201**, 323 (1967).
[4] R.D. Hempstead and M. Lax, Phys. Rev. **161**, 350 (1967).
[5] Y.K. Wang and W.E. Lamb, Jr., Phys. Rev. A **8**, 866, 873 (1973).
[6] S. Grossman and P.H. Richter, Z. Phys. **249**, 43 (1971); **255**, 59 (1972).
[7] M-Tehrani and L. Mandel, Phys. Rev. A **17**, 677 (1978); and Opt. Commun. **16**, 16 (1976).
[8] F.T. Arecchi and A.M. Ricca, Phys. Rev. A **15**, 308 (1977); R. Graham and W.A. Smith, Opt. Commun. **7**, 289 (1973); F. Aronowitz, in *Laser Applications*, edited by M. Ross (Academic, New York, 1971), Vol. I.
[9] See Ref. 1, SSL, p. 119.
[10] See Ref. 1, SSL, p. 133.
[11] See, for example, R.L. Stratonovich, *Topics in Theory of Random Noises* (Gordon and Breach, New York, 1963), Vol. I, p. 75, 102.
[12] See Ref. 11, p. 77.
[13] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. 2, p. 233.
[14] Bateman Manuscript Project, *Tables of Integral Transforms*, edited by A. Erdélyi (McGraw-Hill, New York, 1954), Vol. 1, p. 147 formula (35); and *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. 1, p. 248.
[15] See R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I, p. 292.
[16] See M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, Princeton, N.J., 1965).
[17] F.T. Hioe and E.W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[18] F.T. Hioe, D. MacMillen, and E.W. Montroll, J. Math. Phys. **17**, 1327 (1976).
[19] F.T. Hioe, Phys. Rev. D **15**, 488 (1977); and Phys. Rev. B **16**, 4112 (1977).
[20] E.C. Titchmarsh, *Eigenfunction Expansions* (Oxford University, London, 1962), Part I, p. 151, formula (7.7.6).

On the generation of new solutions of the Einstein–Maxwell field equations from electrovac spacetimes with isometries[a)]

Isidore Hauser and Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 18 November 1977)

We present transformation formulas which facilitate the determination of the metrics, electromagnetic fields, connections and Weyl tensors of those electrovac spacetimes which result when a given solution of the Einstein–Maxwell equations with an isometry is subjected to the transformations of the Kinnersley group. Several applications of our calculational procedures are given as illustrations, and a number of general theorems are presented. In particular, we infer that when we apply such techniques to the only known solution of Petrov type N with twisting principal null rays, the new solutions which result will be algebraically general.

This paper is an exposition concerning *technique*; specifically, technique for constructing new solutions of the Einstein—Maxwell field equations from old "seed solutions." The theory underlying this technique was developed by Geroch[1] for the vacuum case, and later by Kinnersley[2] for the electrovac case, although the discovery of the actual transformations dates back to the pioneer work of Ehlers[3] and Harrison.[4]

The calculational technique which we shall present, and which we shall illustrate with several concrete examples, can best be described as based upon an amalgamation of Geroch's and Kinnersley's formulations of the transformation theory. Thus, on the one hand, we utilize a set of complex vector potentials which reduces in the vacuum case to Geroch's $\{\alpha_i, \beta_i\}$. On the other hand, we also employ the complex scalar potentials $\mathcal{E}$ and Φ introduced by Ernst[5] and utilized by Kinnersley in his development of the transformation theory.

We shall provide formulas for the construction of the transformed connections and Weyl conform tensor as well as the metric and electromagnetic field. Formulas for the Weyl tensor were first developed by Ernst[6] in terms of a specific choice of tetrad, but later it was shown by Ernst and Plebański[7] that such formulas can be derived directly from the Killing vector structural equations and that it is unnecessary to restrict the choice of tetrad in any way. The formulas which we shall present in this paper for the first time are *transformation formulas*, i.e., they relate the new Weyl tensor to the old Weyl tensor. Furthermore, our transformation formulas for the connections are the first such formulas which are valid for an arbitrary tetrad. We believe all these formulas will be found easy to use.

In the last section of this paper several theorems will be given and several observations made concerning the effect of Kinnersley transformations upon seed solutions of certain types.

[a)]Research supported in part by National Science Foundation Grant PHY75-08750.

1. THE KINNERSLEY GROUP

The discovery of the Kinnersley group, first introduced in Ref. 2, did much to unify various transformations which had been developed over a period of many years in order to generate new solutions of the Einstein—Maxwell field equations from old ones. The group can be regarded as a group of nonlinear transformations of certain complex potentials

$$\mathcal{E}^0 := 1, \quad \mathcal{E}^E := \mathcal{E}, \quad \mathcal{E}^M := \Phi, \tag{1.1}$$

where the indices E and M should suggest "Einstein" and "Maxwell," respectively, and the trivial potential $\mathcal{E}^0$ is introduced for purely formal reasons. In terms of these potentials the Kinnersley transformations assume the form

$$\mathcal{E}^{r'} = \Lambda^{-1} B^r{}_s \mathcal{E}^s, \quad \Lambda := B^0{}_t \mathcal{E}^t, \tag{1.2}$$

where the indices r, s, t assume values 0, E, and M. The parameters $B^r{}_s$ are constants governed by the conditions

$$G_{rs}(B^r{}_t)^* B^s{}_u = G_{tu}, \tag{1.3}$$

where

$$\{G_{rs}\} = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.4}$$

Any two matrices $\{B^r{}_s\}$ which differ only by a phase factor will be regarded as equivalent. Thus, the group with which we are concerned is U(2, 1)/U(1), which is isomorphic to SU(2, 1). The transformed gravitational field will be seen to depend only on the parameters

$$b_r := B^0{}_r, \tag{1.5}$$

which therefore play an especially important role.

In Ref. 2 the parametrization of the group was based upon five particular transformations, which correspond to the following matrices:

$$(\mathrm{I}) = \begin{bmatrix} 1 & 0 & 0 \\ -a a^* & 1 & -2a^* \\ a & 0 & 1 \end{bmatrix} \text{ (gauge transformation)},$$

 0022-2488/78/1906-1316$1.00

$$(\mathrm{II})=\begin{bmatrix}1 & 0 & 0\\ i\alpha & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\ \text{(gauge transformation)},$$

$$(\mathrm{III})=\begin{bmatrix}e^{\chi} & 0 & 0\\ 0 & e^{-\chi} & 0\\ 0 & 0 & e^{i\epsilon}\end{bmatrix}\ \text{(uniform conformal mapping; duality rotation)},$$

$$(\mathrm{IV})=\begin{bmatrix}1 & i\beta & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\ \text{(Ehlers transformation)},$$

$$(\mathrm{V})=\begin{bmatrix}1 & -cc^* & -2c^*\\ 0 & 1 & 0\\ 0 & c & 1\end{bmatrix}\ \text{(Harrison transformation)},$$

An alternative parametrization was developed recently by Tanabe,[8] who utilized two special matrices. His matrix U_1 corresponds to the Harrison transformation (V) with $|c|=1$, while his U_0 corresponds to

$$U_0=\begin{bmatrix}a & b & -2c^*e^{-i\alpha}\\ b & a & 2c^*e^{i\alpha}\\ -c & c & \Delta\end{bmatrix},$$

where $a=\frac{1}{2}(e^{i\alpha}+\Delta e^{-i\alpha})$, $b=\frac{1}{2}(e^{i\alpha}-\Delta e^{-i\alpha})$, and $\Delta^2=1+4|c|^2$. It should be noted in passing that when $e^{i\alpha}=1$ the transformation U_0 reduces to the one which was used by Ernst to generate the Kerr–Newman solution from the Kerr solution[9] and the charged version of the Tomimatsu–Sato solution from the uncharged version.[10]

2. TRANSFORMATION OF THE METRIC

The general procedure for deriving the transformed metric will be illustrated using the Reissner–Nordstrom (RN) solution of the Einstein–Maxwell field equations. This seed solution is conveniently described in terms of the orthonormal tetrad

$$e^1=rd\theta,\quad e^2=r\sin\theta d\phi,\quad e^3=S^{-1/2}dr,\quad e^4=S^{1/2}dT,$$

where $S=S(r)=1-2mr^{-1}+|e|^2r^{-2}$, and in terms of the electromagnetic 2-form

$$F=(er^{-2})drdT=(er^{-2})e^3e^4.$$

The first step of our calculational procedure is to identify the "self-dual part" of the electromagnetic 2-form F. Specifically, we evaluate

$$W^M:=2\mathbb{P}F, \tag{2.1}$$

where $\mathbb{P}$ is the projection operator which extracts a self-dual part of the 2-form upon which it operates. In terms of the duality operator employed in the previous work of both authors[11] this projection operator $\mathbb{P}$ is defined so that

$$\mathbb{P}F=\tfrac{1}{2}(F-i^*F). \tag{2.2}$$

Thus, for RN one obtains

$$W^M=(er^{-2})(e^3e^4-ie^1e^2).$$

By the source-free Maxwell equations W^M is always an exact differential. Hence there exists a 1-form A^M such that

$$W^M=dA^M. \tag{2.3}$$

For RN we easily cast W^M into the form

$$W^M=d[-(e/r)dT+ie\cos\theta\, d\phi].$$

In order to apply the transformation (1.2) of the Kinnersley group, one must select a specific nonnull Killing vector field **K** with respect to which the Lie derivative of the 2-form F vanishes. Then it follows immediately that the 1-form

$$K\lrcorner W^M=d\mathcal{E}^M \tag{2.4}$$

is an exact differential of a complex scalar potential $\mathcal{E}^M$. (In previous papers $\mathcal{E}^M$ was denoted by Φ.) In Eq. (2.4) K denotes the covector of **K** while $\lrcorner$ denotes the Grassmann inner product[11] of the 2-form W^M and the 1-form K. If for RN we select our Killing vector to be given by

$$K^a\partial_a=\partial_\phi,$$

then the 1-form K is given by

$$K=r\sin\theta\, e^2.$$

Thus, for RN Eq. (2.4) yields

$$d\mathcal{E}^M=-i(er^{-1})\sin\theta\, e^1=d(ie\cos\theta).$$

Therefore,

$$\Phi=\mathcal{E}^M=ie\cos\theta.$$

We adopt a gauge for the vector potential A^M such that

$$K\lrcorner A^M=\mathcal{E}^M. \tag{2.5}$$

Hence, for RN we may readily identify

$$A^M=\mathcal{E}^Md\phi-(e/r)dT.$$

Turning now to the evaluation of the gravitational potentials, we begin by constructing the 2-form

$$\omega:=\tfrac{1}{2}dK. \tag{2.6}$$

In particular, for RN we obtain the result

$$\omega=-S(r)^{1/2}\sin\theta\, e^2e^3+\cos\theta\, e^1e^2.$$

Now, in general the self-dual part of ω is *not* an exact differential. Nevertheless, it can be shown from the field equations that

$$W^E:=-4\mathbb{P}(\omega+\Phi^*F) \tag{2.7}$$

is always an exact differential. Hence, there exists a 1-form A^E such that

$$W^E=dA^E. \tag{2.8}$$

In the case of RN one finds

$$W^E=2iS(r)^{1/2}\sin\theta\,(e^1e^4-ie^2e^3)$$
$$-2i(1-|e|^2r^{-2})\cos\theta\,(e^3e^4-ie^1e^2),$$

and this may be cast into the form

$$W^E=d[-2iS(r)r\cos\theta\, dT+(|e|^2-r^2)\sin^2\theta\, d\phi].$$

By assumption the 2-form F has vanishing Lie derivative with respect to the Killing vector **K**. One concludes that the 1-form

$$K\lrcorner W^E=d\mathcal{E}^E \tag{2.9}$$

is an exact differential of a complex scalar potential $\mathcal{E}^E$. (In previous papers $\mathcal{E}^E$ was denoted by $\mathcal{E}$). We always choose the additive constant in $\mathcal{E}$ so that

$$\mathrm{Re}\mathcal{E} = f - |\Phi|^2, \tag{2.10}$$

where

$$f = -K \Gamma K. \tag{2.11}$$

In the case of RN one has

$$f = -r^2 \sin^2\theta,$$

and consequently the complex potential is given by

$$\mathcal{E} = \mathcal{E}^E = -r^2 \sin^2\theta - |e|^2 \cos^2\theta.$$

Finally, we adopt a gauge for the vector potential A^E such that

$$K \Gamma A^E = \mathcal{E}^E. \tag{2.12}$$

Thus, for RN

$$A^E = \mathcal{E}^E d\phi - 2iS(r) r \cos\theta \, dT.$$

As a purely formal device we shall introduce an additional trivial 2-form $W^0 = 0$ and an additional trivial complex scalar potential $\mathcal{E}^0 = 1$. It is also convenient to introduce the symbol A^0 for the differential of an affine parameter for the Killing vector field **K**. For example, in the case RN we have $A^0 = d\phi$. In terms of this notation all the following formulas are valid for $r = 0$, E, and M:

$$\underset{\mathbf{K}}{\pounds} W^r = 0 \quad \text{(from the basic assumption that } \underset{\mathbf{K}}{\pounds} F = 0\text{)}, \tag{2.13a}$$

$$dW^r = 0 \quad \text{(from the field equations)}, \tag{2.13b}$$

$$W^r = dA^r \tag{2.13c}$$

$$d(K \Gamma W^r) = 0 \quad \text{[from (2.13a) and (2.13b)]}, \tag{2.13d}$$

$$K \Gamma W^r = d\mathcal{E}^r, \tag{2.13e}$$

$$K \Gamma A^r = \mathcal{E}^r \quad \text{(gauge condition)}. \tag{2.13f}$$

The norm of the Killing vector in the transformed spacetime is easily determined. Noting that Eq. (2.10) can be written in the form

$$f = G_{rs} (\mathcal{E}^r)^* \mathcal{E}^s, \tag{2.14}$$

the transformation formula (1.2) immediately yields the result

$$f \to f' = |\Lambda|^{-2} f. \tag{2.15}$$

For RN

$$\Lambda = b_0 - b_E (r^2 \sin^2\theta + |e|^2 \cos^2\theta) + i b_M e \cos\theta.$$

The determination of the transformed metric is very simple once the covector K' of the Killing vector **K** has been identified. Consider the self-dual 2-form

$$W := -4\mathbb{P}\omega = 2G_{rs} (\mathcal{E}^r)^* W^s \tag{2.16}$$

and the associated 1-form

$$G := K \Gamma W = 2G_{rs} (\mathcal{E}^r)^* d\mathcal{E}^s. \tag{2.17}$$

Comparing Eqs. (2.14) and (2.17), it is immediately clear that

$$\mathrm{Re}G = df. \tag{2.18}$$

Furthermore, for *any* self-dual 2-form W one has the identity

$$W = 2f^{-1} \mathbb{P}[K(K \Gamma W)] \tag{2.19}$$

Therefore, one has the result

$$dK = -2f^{-1} \mathrm{Re}[\mathbb{P}(KG)],$$

which may be written in the useful form

$$d(f^{-1}K) = -f^{-2} * (K \mathrm{Im} G). \tag{2.20}$$

Substituting the transformation (1.2) into Eq. (2.17), one finds that

$$f^{-1}G \to f'^{-1}G' = f^{-1}G - 2\Lambda^{-1} d\Lambda. \tag{2.21}$$

Hence, Eq. (2.20) yields

$$d(f'^{-1}K') = d(|\Lambda|^2 f^{-1}K) + \Lambda^* (b_s W^s) + (b_r W^r)^* \Lambda, \tag{2.22}$$

where we have taken advantage of the relation

$$b_s W^s = 2f^{-1} \mathbb{P}(K d\Lambda), \tag{2.23}$$

which results from another application of the theorem of Eq. (2.19).

Now, since $(W^r)^* W^s = 0$, it can be shown that $(\mathcal{E}^r)^* W^s + (W^r)^* \mathcal{E}^s$ is an exact differential. Thus, there exists a Hermitian matrix of 1-forms M^{rs} such that

$$(\mathcal{E}^r)^* W^s + (W^r)^* \mathcal{E}^s = dM^{rs}. \tag{2.24}$$

We choose the gauge of these vector potentials M^{rs} so that

$$K \Gamma M^{rs} = (\mathcal{E}^r)^* \mathcal{E}^s. \tag{2.25}$$

Thus, quite generally we have

$$M^{0r} = A^r \quad (r = 0, E, M). \tag{2.26}$$

In the case of RN we have $A^0 = d\phi$ and

$$M^{EE} = |\mathcal{E}^E|^2 d\phi,$$

$$M^{EM} = (\mathcal{E}^E)^* \mathcal{E}^M d\phi - e[2(r - 2m) + |e|^2/r - S(r) r \sin^2\theta] dT,$$

$$M^{MM} = |\mathcal{E}^M|^2 d\phi.$$

In terms of the vector potentials M^{rs} it is possible to write the integral of Eq. (2.22) in the form

$$f'^{-1}K' = (b_r)^* b_s [M^{rs} + (\mathcal{E}^r)^* \mathcal{E}^s f^{-1} K] - A^0. \tag{2.27}$$

It should be noted that in the vacuum case M^{0E} and M^{EE} reduce essentially to the 1-forms corresponding to Geroch's fields α_i and β_i, respectively.

For RN Eq. (2.27) yields

$$K' = -f'(d\phi - \tilde{\omega}' dT),$$

where

$$\begin{aligned}\tilde{\omega}' = {} & (b_0^* b_E - b_E^* b_0)[-2iS(r) r \cos\theta] \\ & + (b_0^* b_M + b_M^* b_0)[-e r^{-1}] \\ & + (b_E^* b_M + b_M^* b_E)(-e)[2(r - 2m) + |e|^2/r \\ & - S(r) r \sin^2\theta].\end{aligned}$$

This result constitutes a slight generalization of the result which was obtained earlier by Ernst[12] using $b_0 = 1$, $b_E = \frac{1}{4}\beta_0^2$, and $b_M = \beta_0$, where $\beta_0 =$ real constant.

Finally, one obtains the transformed metric tensor, or, as was suggested in Ref. 7, one specifies a convenient tetrad for the transformed spacetime constructed from the tetrad which was used to describe the

original spacetime. In terms of our present notation such a tetrad is provided by

$$e^{a\prime} := |\Lambda|[e^a + f^{-1}K(K\lrcorner e^a)] - |\Lambda|^{-1}f'^{-1}K'(K\lrcorner e^a). \quad (2.28)$$

In the case of RN this yields

$$e^{a'} = |\Lambda| e^a \quad (a=1,3,4),$$

$$e^{2'} = |\Lambda|^{-1} r\sin\theta(d\phi - \tilde{\omega}'dT),$$

which agrees with the result given in Ref. 11.

Before we conclude this section, we should like to emphasize that $\{e_a\}$ need not be an orthonormal tetrad. It might be, for example, a *null tetrad* $\{k, m, t, t^*\}$ subject to our convention[13] that the only nonvanishing inner products are given by

$$k\lrcorner m = t\lrcorner t^* = +1. \quad (2.29)$$

As a concrete example consider the null tetrad

$$k = d\rho, \quad m = dv, \quad t = (2)^{-1/2}\rho(d\xi + id\eta),$$

which describes Minkowski space (MS) in terms of coordinates $\{\rho, v, \xi, \eta\}$ such that

$$\rho = (2)^{-1/2}(z - T), \quad v = (2)^{-1/2}(z + T),$$

$$\xi = \rho^{-1}x, \quad \eta = \rho^{-1}y.$$

From among all the Killing vectors which Minkowski space possesses let us select

$$K^a\partial_a = \partial_\eta.$$

Then the methods which we have described in this section can be invoked in order to show that the new tetrad given by Eq. (2.28) has the form

$$k' = |\Lambda| d\rho, \quad m' = |\Lambda| dv,$$

$$t' = (2)^{-1/2}\{|\Lambda| d\xi + i|\Lambda|^{-1} \times [d\eta + 2i(b_0{}^*b_E - b_E{}^*b_0)\xi\rho d\rho]\},$$

where

$$\Lambda = b_0 - b_E\rho^2.$$

In the MS example

$$W^M = 0, \quad W^E = -4\mathbb{P}\omega = 2(2)^{1/2}ikt,$$

and the complex scalar potentials are given by

$$\mathcal{E}^M = 0, \quad \mathcal{E}^E = f = -\rho^2.$$

3. TRANSFORMATION OF THE ELECTROMAGNETIC FIELD

Now we shall consider the effect of the transformation (1.2) upon the Cartan components of the electric and magnetic fields. For this purpose we construct from the tetrad $\{e^a\}$ a basis $\{B_A\}$ for self-dual 2-forms. If $\{e^a\}$ is an *orthonormal tetrad*, then a convenient basis $\{B_A\}$ is defined by

$$B_1 = e^1e^4 - ie^2e^3,$$

$$B_2 = e^2e^4 - ie^3e^1, \quad (3.1)$$

$$B_3 = e^3e^4 - ie^1e^2.$$

With this choice of basis the components $F_A := F\lrcorner B_A$ of the electromagnetic 2-form F are related to the Cartan components of the electric field E_A and the magnetic field H_A by

$$F_A = -(E_A + iH_A). \quad (3.2)$$

In the case RN the components of the electromagnetic 2-form F are given by

$$F_1 = F_2 = 0, \quad F_3 = -er^{-2},$$

while the components $\omega_A := \omega\lrcorner B_A$ of the Killing 2-form ω are given by

$$\omega_1 = -iS(r)^{1/2}\sin\theta, \quad \omega_2 = 0, \quad \omega_3 = i\cos\theta.$$

We shall demonstrate later that the latter quantities play a key role in the determination of the transformed Weyl conform tensor.

From the tetrad (2.28) describing the transformed metric we can construct a basis $\{B_A'\}$ with respect to which we can define components F_A' and ω_A' of the transformed 2-forms F' and ω'. As a result of the transformation (1.2) the components of F and ω transform as follows:

$$F_A' = \Lambda^{-2}(\det B)[\Lambda^*(1 - 2b_Ef\Lambda^{-1})^*F_A + (b_M - 2b_E\mathcal{E}^{M*})^*\omega_A], \quad (3.3)$$

$$\omega_A' = |\Lambda|^{-2}[(1 - 2b_Ef\Lambda^{-1})\omega_A + f\Lambda^{-1}(b_M - 2b_E\mathcal{E}^{M*})F_A]. \quad (3.4)$$

These relations are derived by observing that according to Eq. (2.19)

$$W = 2f^{-1}\mathbb{P}K(K\lrcorner W)]$$

for any self-dual 2-form W, and in particular for $\mathbb{P}F$ and $\mathbb{P}\omega$. This implies, however, that

$$W_A = -2f^{-1}(K\lrcorner W)\lrcorner(K\lrcorner B_A). \quad (3.5)$$

On the other hand, careful consideration of the transformation (2.28) reveals that

$$W_A' = -2f^{-1}(K\lrcorner W)'\lrcorner(K\lrcorner B_A), \quad (3.6)$$

where $(K\lrcorner W)' = K'\lrcorner' W'$, $\lrcorner'$ being the Grassmann inner product with respect to the transformed metric. We then use the transformation properties of

$$K\lrcorner\mathbb{P}F = \tfrac{1}{2}d\mathcal{E}^M$$

and

$$K\lrcorner\mathbb{P}\omega = -\tfrac{1}{4}G.$$

which are already known, in order to infer relations (3.3) and (3.4). However, it should be noted that the identity

$$B^r{}_sB^w{}_t - B^r{}_tB^w{}_s = (\det B)\delta^{rxw}_{stu}G^{uv}G_{xy}(B^y{}_v)^* \quad (3.7)$$

has been used to simplify Eq. (3.3).

The parameters b_0, b_E and b_M are *not* independent complex parameters, but rather they are subject to the constraint

$$G^{rs}b_r{}^*b_s = 0. \quad (3.8)$$

Furthermore, $\det B$ is necessarily of modulus unity.

In the RN case Eq. (3.3) yields a generalization of the results obtained in Ref. 12, where the particular values $b_0 = 1$, $b_E = \frac{1}{4}\beta_0{}^2$, $b_M = \beta_0$, and $\det B = 1$ were selected in order to end up with a solution with certain desired properties. (Due to different conventions the

sign of our potential $\mathcal{E}^M = \Phi$ differs from that of the Φ found in Ref. 12, but both the metric and the electric and magnetic fields agree.)

If $\{e_a\}$ is a *null tetrad* $\{k, m, l, t^*\}$ then a convenient basis $\{B_A\}$ for self-dual 2-forms is provided by

$$B_{+1} = kl, \quad B_0 = km + lt^*, \quad B_{-1} = mt^*. \tag{3.9}$$

In the Minkowski space example the initial electromagnetic field vanishes, while

$$\omega_{+1} = \omega_0 = 0, \quad \omega_{-1} = -(2)^{-1/2} i.$$

Furthermore, Eqs. (3.3) and (3.4) reduce to

$$F_A' = \Lambda^{-2}(\det B) b_M{}^* \omega_A$$

and

$$\omega_A' = |\Lambda|^{-2}(1 - 2b_E f \Lambda^{-1})\omega_A,$$

respectively.

4. TRANSFORMATION OF THE CONNECTIONS

We shall now introduce a triplet of connection 1-forms

$$v_A := \tfrac{1}{2} B_A{}^{ab} d\mathbf{e}_a \rceil \mathbf{e}_b, \tag{4.1}$$

where

$$B_A{}^{ab} := (e^a e^b) \sqcap B_A. \tag{4.2}$$

The components of v_A with respect to the tetrad $\{e_b\}$ will be denoted by

$$v_{Ab} := v_A \sqcap e_b. \tag{4.3}$$

In the particular case of a null tetrad both of the authors have used the special symbols

$$v := v_{+1}, \quad u := v_0, \quad w := v_{-1}, \tag{4.4}$$

in their previous publications. The components of these three 1-forms they have designated by v_k, v_m, v_t, v_{t^*}, etc. We shall continue to employ this specialized notation wherever it seems appropriate. The reader is advised to consult Ref. 13 for correspondences with other notations.

In the Minkowski space example considered in the previous sections the connection forms v and u vanish, while

$$w = v_{-1} = \rho^{-1} t^*.$$

Thus, 11 of the 12 components of v, u, and w vanish, the only nonvanishing component being

$$w_t = \rho^{-1}.$$

Since we have already constructed the transformed metric, we could evaluate the transformed connections directly. However, there are several advantages to be gained by developing a transformation formula for the connections. If the evaluation of the connection 1-forms was difficult for the seed metric, the evaluation for the transformed metric is not likely to be easier. Furthermore, there may be instances where the determination of the transformed metric is difficult, while the determination of the transformed connections is relatively simple. Bearing this in mind, we now present a formula for transforming the components of the connection 1-forms; namely,

$$v_{Ab}' = |\Lambda|^{-1}[v_{Ab} + \Lambda^{-1} d\Lambda \sqcap (e_b \sqcap B_A) - K_b \Lambda^{-1} b_s W_A^s], \tag{4.5}$$

where $K_b := K \sqcap e_b$. This has been proved only for tetrads such that $\underset{\mathbf{K}}{\pounds} e^a = 0$.

In the MS example only $w_t = v_{-1t} \neq 0$ and only $W^E{}_{-1} \neq 0$. Furthermore,

$$K = \rho^2 d\eta = -(2)^{-1/2} i\rho(l - l^*),$$

so

$$K_k = K_m = 0, \quad K_t = (2)^{-1/2} i\rho, \quad K_{t^*} = -(2)^{-1/2} i\rho.$$

Since $\Lambda = b_0 - b_E \rho^2$, we see that

$$d\Lambda = -2b_E \rho k.$$

Thus, no term in Eq. (4.5) contributes to $v_{+1b}' = v_b'$, while only the $d\Lambda$ term contributes to $v_{0b}' = u_b'$, and then only for $e_b = m$. The values of the transformed connections are given by

$$v_k' = v_m' = v_t' = v_{t^*}' = 0,$$

$$u_k' = u_t' = u_{t^*}' = 0, \quad u_m' = 2b_E |\Lambda|^{-1} \Lambda^{-1} \rho,$$

$$w_k' = w_m' = 0, \quad w_t' = |\Lambda|^{-1} \rho^{-1}, \quad w_{t^*}' = -2b_E |\Lambda|^{-1} \Lambda^{-1} \rho.$$

It is already apparent that the transformed metrics are at most of Petrov type N, since $v' = 0$, while

$$u' = 2b_E \Lambda^{-1} \rho d\rho$$

is an exact differential. The solutions which the Kinnersley group has generated here are in fact Petrov type N plane fronted waves.[14]

Sometimes it may be convenient to employ a tetrad $\{e^a\}$ which does not satisfy the criterion $\underset{\mathbf{K}}{\pounds} e^a = 0$. While Eq. (4.5) may not be valid, one may use instead the following generally valid transformation law for the connection 1-forms v_A themselves:

$$(v_A + f^{-1} K \omega_A)' = (v_A + f^{-1} K \omega_A) - \Lambda^{-1}\{d\Lambda \sqcap B_A - f^{-1} K[d\Lambda \sqcap (K \sqcap B_A)]\}. \tag{4.6}$$

The transformed quantities $f'^{-1} K'$ and ω_A' were evaluated in Secs. 2 and 3 of this paper.

In addition to being useful when Eq. (4.5) does not hold, Eq. (4.6) facilitates the derivation of the transformation law for the Weyl conform tensor, which we consider in the next section.

5. TRANSFORMATION OF THE WEYL CONFORM TENSOR

In Ref. 6 formulas for null tetrad components of the Weyl conform tensor were given. However, these formulas involved a special choice for the null tetrad, and in Ref. 7 it was shown that such formulas follow directly from the Killing vector structural equations, with no restriction upon the null tetrad being necessary. We shall develop here analogous formulas for an arbitrary tetrad, employing the language of differential forms rather than that of spinors. From these formulas we shall derive the transformation properties of the Weyl tensor components under the Kinnersley group. The resulting transformation law will be sufficiently simple that one may infer from that law some

important attributes of the transformed spacetime even before the metrics are evaluated.

If ω denotes the covector of the differential form $\omega = \frac{1}{2}dK$, the Killing vector structural equations can be written in the form

$$d\mathbf{K} = \omega \rceil \mathbb{1}, \tag{5.1a}$$

$$d\omega = K \rfloor \mathbb{R}. \tag{5.1b}$$

where $\mathbb{1} := e^a \mathbf{e}_a$ is the unit (1, 1)-form and

$$\mathbb{R} := \tfrac{1}{2}\mathbf{e}^a \wedge d^2\mathbf{e}_a = \tfrac{1}{4}R_{ab}{}^{cd} e^a e^b \mathbf{e}_c \wedge \mathbf{e}_d \tag{5.2}$$

is the invariant Riemann (2, 2)-form.

The irreducible parts of the Riemann tensor correspond to the quantities

$$R := 4 G_{AB} B^A \rfloor \mathbb{R} \rceil \mathbf{B}^B, \tag{5.3a}$$

$$R^{\bar{A}B} := (B^A)^* \rfloor \mathbb{R} \rceil \mathbf{B}^B, \tag{5.3b}$$

$$C^{AB} := B^A \rfloor \mathbb{R} \rceil \mathbf{B}^B - \tfrac{1}{12} R G^{AB}, \tag{5.3c}$$

where

$$G^{AB} := B^A \rfloor B^B, \tag{5.4}$$

and $\mathbf{B}^A$ is the covector of B^A.

We evaluate the Weyl tensor components C^{AB} by using the theorem of Eq. (2.19) once again. Thus,

$$C^{AB} + \tfrac{1}{12} R G^{AB} = -2 f^{-1} (K \rfloor B^A) \rfloor [K \rfloor (\mathbb{P}\mathbb{R}) \rceil \mathbf{B}^B]$$
$$= -2 f^{-1} (K \rfloor B^A) \rfloor [K \rfloor (\mathbb{R} - \mathbb{P}^* \mathbb{R}) \rceil \mathbf{B}^B].$$

From Eqs. (5.1b) and (5.3c) we conclude that

$$C^{AB} + \tfrac{1}{12} R G^{AB} = -2 f^{-1} (K \rfloor B^A) \rfloor [d\omega \rceil \mathbf{B}^B - K \rfloor (B_C)^* R^{\bar{C}B}].$$

However, using the relation

$$d\boldsymbol{\omega} \rceil \mathbf{B}^B = d(\boldsymbol{\omega} \rceil \mathbf{B}^B) - \boldsymbol{\omega} \rceil d\mathbf{B}^B = d\omega^B + \epsilon^{BCD} v_C \omega_D,$$

we obtain

$$C^{AB} + \tfrac{1}{12} R G^{AB} = -2 f^{-1} (K \rfloor B^A) \rfloor [d\omega^B + \epsilon^{BCD} v_C \omega_D - K \rfloor (B_C)^* R^{\bar{C}B}]. \tag{5.5}$$

It should be noted that whenever the authors have used a null tetrad, they have denoted C^{AB} by C_{-A-B}, e.g., C^{11} by C_{-2}. The reader should consult Ref. 13 for the correspondences with other notations in common use.

In the electrovac case, where the transformations of the Kinnersley group apply, we have

$$R^{\bar{A}B} = -2(F^A)^* F^B, \quad R = 0, \tag{5.6}$$

and Eq. (5.5) reduces to

$$C^{AB} = -2 f^{-1} (K \rfloor B^A) \rfloor [d\omega^B + \epsilon^{BCD} v_C \omega_D + 2(K \rfloor B_C{}^*)(F^C)^* F^B]. \tag{5.7}$$

Careful consideration of the transformation (2.28) reveals that the Weyl tensor components of the transformed metric are given by

$$C^{AB'} = -2 f^{-1} (K \rfloor B^A) \rfloor [d\omega^{B'} + \epsilon^{BCD} v_C{}' \omega_D' + 2(K \rfloor B_C{}^*)(F^{C'})^* F^{B'}]. \tag{5.8}$$

The substitution of Eqs. (3.4) and (4.6) into this expression for $C^{AB'}$ yields the transformation law for the Weyl tensor components.

At the present time we shall present only the form of the transformation law that is obtained *when the initial metric is a vacuum spacetime*; namely,

$$C^{AB'} = |\Lambda|^{-2} (1 - 2 b_E f \Lambda^{-1}) \{C^{AB} + 12 b_E \Lambda^{-1} [\omega^A \omega^B - \tfrac{1}{3}(\omega^C \omega_C) G^{AB}]\}. \tag{5.9}$$

If the initial spacetime is Minkowski space, i.e., if $C^{AB} = 0$, then Eq. (5.9) tells us that the transformed space can be at most of type N or type D *regardless of which Killing vector* K *of Minkowski space is selected.* This conclusion generalizes our findings concerning the particular MS example we considered earlier. In the next section we shall describe several other simple theorems that may be of considerable practical use.

6. RECAPITULATION AND CONCLUSION

We have provided here a systematic calculational procedure for the generation of solutions of the Einstein–Maxwell field equations. The section titles should facilitate the location of any desired formula.

One begins, of course, by choosing any convenient tetrad, e.g., an orthonormal tetrad or a null tetrad. The self-dual 2-forms W^r are determined by Eqs. (2.1) and (2.7), and then Eqs. (2.13a)–(2.13f) are employed to identify the 1-forms A^r and the 0-forms $\mathcal{E}^r$. Finally, the 1-forms M^{rs} are determined using Eqs. (2.24)–(2.26).

The effect of the Kinnersley transformations upon f and K is ascertained using Eqs. (2.15) and (2.27), respectively. Then the tetrad for the transformed spacetime is constructed using Eq. (2.28). This is equivalent to specifying the transformed metric tensor.

The transformed electromagnetic field is ascertained using Eq. (3.3) after a convenient basis for self-dual 2-forms has been selected. Simultaneously one may work out the transformed Killing 2-form components using Eq. (3.4), noting that these quantities play an important role in the determination of the Weyl tensor.

If the tetrad $\{e^a\}$ satisfies the condition ${}_K\!\pounds e^a = 0$, then the components of the connection 1-forms transform according to Eq. (4.5). Otherwise, one can always employ Eq. (4.6) to determine the transformed connection 1-forms. Sometimes the Petrov type of the transformed spacetime is immediately obvious from the transformed connections. In any event the components of the Weyl tensor can be evaluated using Eq. (5.8), or one may use Eq. (5.9) if the seed solution is a vacuum field.

The effort which we have devoted to the development of an efficient calculational technique for generating new solutions of the Einstein–Maxwell field equations from old ones was in large part motivated by our determination to see what new solutions would be spawned by Hauser's first TNT solution.[15] We have in fact initiated such an endeavor by identifying the complex potential $\mathcal{E}$ of this seed solution.[16]

Although we have only just begun applying the techni-

ques of this paper to the TNT solution, we can already infer that the solutions we shall generate will be of Petrov type I (i.e., algebraically general), for consideration of Eq. (5.9) shows that the following theorem holds:

Theorem 1: From a Petrov type N vacuum field other than a plane fronted gravitational wave only Petrov type I fields are generated (except for the special case $b_E = b_M = 0$, where a type N field differing trivially from the seed solution is generated).

It will be particulary interesting to see if any of the solutions generated from Hauser's TNT solution possesses more than one Killing vector, for such a solution with multiple Killing vectors would serve as a seed solution for the generation of yet other solutions of the Einstein–Maxwell field equations. On the other hand, if there exists no such solution with multiple Killing vectors, this would constitute the first proof that there exist solutions which cannot ultimately be connected to Minkowski space by a sequence of Kinnersley transformations.

It should be clear that further consideration of Eqs. (5.8) and (5.9) should yield theorems concerning seed solutions of other Petrov types. It is also possible to infer useful theorems from Eqs. (4.5) and (4.6). Thus, for example, we have the following:

Theorem 2: Suppose $\mathbf{k}$ is a null geodesic vector field in the original spacetime, and ${}_{\mathbf{K}}\pounds k = 0$, $K \lrcorner k = 0$. Then $\mathbf{k}' = |\Lambda|\mathbf{k}$ is a null geodesic field in the transformed spacetime.

Theorem 3: Suppose $\mathbf{k}$ is a shear-free null geodesic field in the original spacetime, and $\mathbf{k}$ is an eigenvector of both ω and F. Then

$$\mathbf{k}' = |\Lambda|[\mathbf{k} + f^{-1}\mathbf{K}(K \lrcorner k)] - |\Lambda|^{-1} f'^{-1}\mathbf{K}'(K \lrcorner k)$$

is a shear-free null geodesic field in the transformed spacetime, and $\mathbf{k}'$ is an eigenvector of both ω' and F'.

Corollary: Suppose $\mathbf{k}$ and $\mathbf{m}$ are the principal null vector fields of a type D seed solution, and $\mathbf{k}$ and $\mathbf{m}$ are eigenvectors of both ω and F. Then the transformed spacetime is also type D, and $\mathbf{k}'$ and $\mathbf{m}'$ are eigenvectors of both ω' and F'.

APPENDIX: GRASSMANN INNER PRODUCTS

The tangent manifold at a point of a spacetime is a four-dimensional linear vector space with an inner product. If four linearly independent vectors are denoted by $\mathbf{e}_\alpha$ ($\alpha = 1, 2, 3, 4$), then the matrix $g_{\alpha\beta} = \mathbf{e}_\alpha \cdot \mathbf{e}_\beta$ is nonsingular and has signature +2.

Any tangent vector can be expanded in terms of the basis $\mathbf{e}_\alpha$. The covector u of $\mathbf{u} = u^\alpha \mathbf{e}_\alpha$ is the differential form $u = u^\alpha e_\alpha$, where e_α is a linear functional on tangent vectors which has the value $g_{\alpha\beta}$ when it acts upon $\mathbf{e}_\beta$.

A basis for p-vectors is provided by

$$\mathbf{e}_{(p|\alpha)} = \mathbf{e}_{\alpha_1} \wedge \mathbf{e}_{\alpha_2} \wedge \cdots \wedge \mathbf{e}_{\alpha_p} \quad (\alpha_1 < \alpha_2 < \cdots < \alpha_p)$$

and a basis for p-forms is provided by

$$e_{(p|\alpha)} = e_{\alpha_1} e_{\alpha_2} \cdots e_{\alpha_p} \quad (\alpha_1 < \alpha_2 < \cdots < \alpha_p),$$

where we suppress the wedge symbol between differential forms when an exterior product is to be understood.

The ordinary inner product of a p-vector

$$\mathbf{u} = \frac{1}{p!} u^{\alpha_1 \cdots \alpha_p} \mathbf{e}_{\alpha_1} \wedge \cdots \wedge \mathbf{e}_{\alpha_p} = u^{(p|\alpha)} \mathbf{e}_{(p|\alpha)}$$

and a p-vector

$$\mathbf{v} = \frac{1}{p!} v^{\alpha_1 \cdots \alpha_p} \mathbf{e}_{\alpha_1} \wedge \cdots \wedge \mathbf{e}_{\alpha_p} = v^{(p|\alpha)} \mathbf{e}_{(p|\alpha)}$$

is defined by

$$\mathbf{u} \cdot \mathbf{v} := \frac{1}{p!} u^{\alpha_1 \cdots \alpha_p} v_{\alpha_1 \cdots \alpha_p} = u^{(p|\alpha)} v_{(p|\alpha)},$$

where scripts are lowered using the metric $g_{\alpha\beta} = \mathbf{e}_\alpha \cdot \mathbf{e}_\beta$, and where $(p|\alpha)$ is a shorthand notation for the row of scripts $\alpha_1, \alpha_2, \ldots, \alpha_p$ subject to the condition $\alpha_1 < \alpha_2 < \cdots < \alpha_p$. The Grassmann inner product $\mathbf{u} \lrcorner \mathbf{v}$ of a p-vector $\mathbf{u}$ and a q-vector $\mathbf{v}$ is defined by

$$\mathbf{u} \lrcorner \mathbf{v} = 0 \quad \text{if } p < q,$$

$$\mathbf{u} \lrcorner \mathbf{v} = \mathbf{u} \cdot \mathbf{v} \quad \text{if } p = q,$$

and if $p > q$ it is the $(p\text{-}q)$-vector such that

$$(\mathbf{u} \lrcorner \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot (\mathbf{v} \wedge \mathbf{w})$$

for all $(p\text{-}q)$-vectors $\mathbf{w}$.

The following properties of the Grassmann inner product can be derived:

(1) $\mathbf{u} \lrcorner \mathbf{v}$ is bilinear in $\mathbf{u}$ and $\mathbf{v}$.

(2) $(\mathbf{u} \lrcorner \mathbf{v}) \lrcorner \mathbf{w} = \mathbf{u} \lrcorner (\mathbf{v} \wedge \mathbf{w})$ for all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$.

(3) $(\mathbf{u} \wedge \mathbf{v}) \lrcorner \mathbf{w} = (\mathbf{u} \lrcorner \mathbf{w})\mathbf{v} - \mathbf{u} \wedge (\mathbf{v} \lrcorner \mathbf{w})$ for any p-vector $\mathbf{v}$ and any 1-vectors $\mathbf{u}$ and $\mathbf{w}$.

These properties permit the simple evaluation of the Grassmann inner products encountered in practice.

The Grassmann inner product $u \lrcorner v$ of a p-form u and a q-form v is defined by

$$u \lrcorner v = 0 \quad \text{if } p > q,$$

$$u \lrcorner v = u \cdot v = \mathbf{u} \cdot \mathbf{v} \quad \text{if } p = q,$$

and if $p < q$ it is the $(q\text{-}p)$-form such that

$$w \cdot (u \lrcorner v) = (wu) \cdot v$$

for all $(q\text{-}p)$-forms w.

The following properties of this Grassmann inner products can be derived:

(1) $u \lrcorner v$ is bilinear in u and v.

(2) $w \lrcorner (u \lrcorner v) = (wu) \lrcorner v$ for all forms $u, v, w,$.

(3) $w \lrcorner (uv) = u(w \lrcorner v) - (w \lrcorner u)v$ for any p-forms u and any 1-forms v and w.

Again these properties permit the simple evaluation of the Grassmann inner products encountered in practice.

A (p, q)-form

$$\mathbf{A} = A_{(p|\alpha)}{}^{(q|\beta)} e^{(p|\alpha)} \mathbf{e}_{(q|\beta)}$$

is a $\binom{q}{p}$ tensor such that $\mathbf{A}_{\alpha_1 \cdots \alpha_p}{}^{\beta_1 \cdots \beta_q} = A_{[\alpha_1 \cdots \alpha_p]}{}^{[\beta_1 \cdots \beta_q]}$.

If $\mathbf{A}$ is a (p,q)-form and $\mathbb{B}$ is an (r,s)-form, then we shall define

$$\mathbf{A}\wedge\mathbb{B}=A_{(p|\alpha)}{}^{(q|\beta)}B_{(r|\gamma)}{}^{(s|\delta)}[e^{(p|\alpha)}e^{(r|\gamma)}][\mathbf{e}_{(q|\beta)}\wedge\mathbf{e}_{(s|\delta)}],$$

$$\mathbf{A}\lrcorner\mathbb{B}=A_{(p|\alpha)}{}^{(q|\beta)}B_{(r|\gamma)}{}^{(s|\delta)}[e^{(p|\alpha)}e^{(r|\gamma)}][\mathbf{e}_{(q|\beta)}\lrcorner\mathbf{e}_{(s|\delta)}],$$

$$\mathbf{A}\llcorner\mathbb{B}=A_{(p|\alpha)}{}^{(q|\beta)}B_{(r|\gamma)}{}^{(s|\delta)}[e^{(p|\alpha)}\llcorner e^{(r|\gamma)}][\mathbf{e}_{(q|\beta)}\wedge\mathbf{e}_{(s|\delta)}].$$

The general definitions of the Grassmann inner products are set up so that

$$(\mathbf{A}\lrcorner\mathbb{B})\lrcorner\mathbb{C}=\mathbf{A}\lrcorner(\mathbb{B}\wedge\mathbb{C}),$$

$$\mathbf{A}\llcorner(\mathbb{B}\llcorner\mathbb{C})=(\mathbf{A}\wedge\mathbb{B})\llcorner\mathbb{C},$$

regardless of the degrees of $\mathbf{A}$, $\mathbb{B}$, and $\mathbb{C}$.

The Grassmann inner products[17] are useful in connection with the evaluation of the duals of forms and vectors. If u is a p-vector, then

$$*\mathbf{u}=-\mathbf{e}^{1234}\lrcorner\mathbf{u},$$

and if u is a p-form, then

$$*u=-(-1)^{p}u\llcorner e^{1234}.$$

Here $e^{1234}:=e^1e^2e^3e^4$ is the volume element constructed from an orthonormal basis $\{e^a\}$.

[1]R. Geroch, J. Math. Phys. **12**, 918 (1971).
[2]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[3]J. Ehlers, *Les théories relativistes de la gravitation* (CNRS, Paris, 1959).
[4]B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[5]F. J. Ernst, Phys. Rev. **167**, 1175 (1968); **168**, 1415 (1968); J. Math. Phys. **15**, 1409 (1974).
[6]F. J. Ernst, J. Math. Phys. **15**, 2027 (1974).
[7]F. J. Ernst and J. F. Plebański, Ann. Phys. (N.Y.) **107**, 266 (1977).
[8]Y. Tanabe, preprint.
[9]F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[10]F. J. Ernst, Phys. Rev. D **7**, 2520 (1973).
[11]See the appendix of the present paper.
[12]F. J. Ernst, J. Math. Phys. **17**, 54 (1976).
[13]F. J. Ernst, J. Math. Phys. **19**, 489 (1978).
[14]W. Kundt, Z. Phys. **163**, 77 (1961).
[15]I. Hauser, Phys. Rev. Lett. **33**, 1112 (1974); J. Math. Phys. **19**, 661 (1978).
[16]The MS example introduced in Sec. 2 of this paper corresponds to the "flat space limit" of Hauser's TNT solution.
[17]I. Hauser, Proceedings of the Relativity Seminar, PORS-IIT-4, Illinois Institute of Technology, may be consulted for additional information concerning the Grassmann inner products.

Approximate symmetry groups of inhomogeneous metrics: Examples

Alan Spero

Department of Physics, Wesleyan University, Middletown, Connecticut 06457
and Department of Physics, Northern Michigan University, Marquette, Michigan 49855 [a]

Ralph Baierlein

Department of Physics, Wesleyan University, Middletown, Connecticut 06457
(Received 27 October 1977)

By definition, an N-dimensional positive-definite inhomogeneous metric is not invariant under any N-parameter, simply-transitive continuous group of motions. Nonetheless, it is possible to construct a group (simply-transitive and of N parameters) that comes closest to leaving the given metric invariant. We call this group the approximate symmetry group of the metric. In an earlier paper, we described a technique for constructing the approximate symmetry group of a given metric. Here, we briefly review that technique and then present some examples of its application. All two-dimensional metrics are analyzed, and simple criteria are given for determining their approximate symmetry groups. Three three-dimensional metrics are investigated: the invariant hypersurfaces of the Kantowski–Sachs space-times and two families of hypersurfaces in the Gowdy T^3 space-times. The approximate symmetry group of the former is found to be of Bianchi Type I and those of the latter may be I or VI_0. Defining, via our technique, a measure I of the magnitude of a metric's inhomogeneity, we study the time dependence of I for the hypersurfaces in the Gowdy metric. We find it is possible in some cases for these hypersurfaces to approach homogeneity ($I \to 0$) both in the asymptotic future and near the initial singularity. Finally, we constant a homogeneous background metric for these hypersurfaces.

I. INTRODUCTION

In a previous paper,[1] we proposed a precise definition for the notion of an approximate symmetry of an inhomogeneous, positive-definite, Riemannian metric, together with a technique for determining the simply-transitive Lie group describing such an approximate symmetry. Succinctly, given a metric we use the calculus of variations to search for a complete set of orthonormal vector fields whose commutation coefficents cients differ as little as possible from some set of structure constants. These vectors then approximate the vector generators of a simply-transitive Lie group. The group reciprocal to this group may be considered to be the approximate symmetry group of the metric. The metric can be classified according to its approximate symmetry group by using the Bianchi classification of Lie algebras.[2]

Herein, we present examples of the technique's application. After reviewing the essential details of our technique (Sec. II), we apply it to two-dimensional spaces (Sec. III). There, the computational difficulties disappear, and we are able to give criteria permitting easy classification of all two-dimensional spaces. Two families of inhomogeneous spacelike hypersurfaces in the Gowdy T^3 spacetimes[3] are studied in Sec. IV. These enable us to examine the behavior of our classification system in an evolving spacetime. In Sec. V, we construct a spatially homogeneous background metric for the Gowdy hypersurfaces. This background metric is invariant under motions induced by the approximate symmetry group of certain spacelike hypersurfaces in the Gowdy metric. In Sec. VI, we discuss the results.

Finally, in Appendix A we analyze the metrics belonging to a family of spacelike hypersurfaces in the Kantowski–Sachs spacetimes.[4] This example is included, in part, for completeness: The Kantowski–Sachs spacetimes are the only ones which may be considered spatially homogeneous, but do not meet our technical definition of homogeneity.

II. APPROXIMATE SYMMETRY GROUPS

Given a subset $\mathcal{U}$ of an N-dimensional surface $\mathcal{S}$ with positive-definite metric g, we wish to find a set of N orthonormal vectors[5] $\{\mathbf{e}_A\}$, $A = 1, \ldots, N$, in $\mathcal{U}$ such that their commutation coefficients

$$\gamma^C_{AB} = g_{ab}[\mathbf{e}_A, \mathbf{e}_B]^a e^{Cb} = 2e^a_{[A}\nabla_a e^b_{B]} e^C_b \tag{2.1}$$

are as close as possible to some set of structure constants C^C_{AB}. To do this, we require that $\{\mathbf{e}_A\}$ and C^C_{AB} give a global minimum of

$$I \equiv \frac{1}{V}\int_{\mathcal{U}} \Delta^C_{AB}\Delta^C_{AB}\, dV + 8\lambda_A n^{AB} a_B + \lambda_{[AB]}(n^{AB} - n^{BA}), \tag{2.2}$$

where

$$\Delta^C_{AB} \equiv \gamma^C_{AB} - C^C_{AB} = \gamma^C_{AB} - \{\epsilon_{ABD} n^{DC} + 2a_{[A}\delta^C_{B]}\} \tag{2.3}$$

and

$$V = \int_{\mathcal{U}} dV = \int_{\mathcal{U}} (\det g)^{1/2} dx^1 \cdots dx^N. \tag{2.4}$$

The Lagrange multipliers λ_A and $\lambda_{[AB]}$ are included to insure that C^C_{AB} (or, equivalently, n^{AB} and a_A) satisfy the Jacobi condition:

$$n^{AB} a_B = 0 \tag{2.5a}$$

and

[a] Present address.

 0022-2488/78/1906-1324$1.00

$$n^{AB} = n^{BA}. \tag{2.5b}$$

For a homogeneous metric, that is, and N-dimensional, positive-definite metric invariant under a simply-transitive, N-parameter Lie group, there will be a set of vectors (invariant under the group) whose γ^C_{AB} are exactly structure constants. For these metrics, $I=0$ at its minimum. For all other metrics, $I>0$ at its minimum. We may consider I to be a measure of the metric's inhomogeneity.

The nessary condition for a minimum is

$$\delta I = 0, \tag{2.6}$$

where the variation is with respect to the $\mathbf{e}_A$ (subject to the orthonormality constraint), n^{AB}, a_A, λ_A, and $\lambda_{[AB]}$. After one performs the variation, a bit of algebra leads to three sets of equations:

(a) a set of N coupled parital differential equations,

$$\nabla\cdot[\mathbf{e}^A\Delta^C_{AB}]\epsilon^{BCM} + \Lambda^{CM}_{AB}\Delta^C_{AB} = 0, \tag{2.7}$$

where

$$\Lambda^{CM}_{AB} = \Lambda^{CM}_{[AB]} \equiv \tfrac{1}{2}[C^D_{AB}\epsilon^{DCM} + C^C_{AD}\epsilon^{DBM} - C^C_{BD}\epsilon^{DAM}]; \tag{2.8}$$

(b) the Jacobi conditions (2.5), together with a set of algebraic equations connecting n^{AB} and a_A with the average of γ^C_{AB}:

$$n^{AB} = \bar{\eta}^{AB} - (\lambda^A a^B + a^A\lambda^B), \tag{2.9}$$

$$a_A = \bar{\alpha}_A - n_{AB}\lambda^B; \tag{2.10}$$

here, $\bar{\eta}^{AB}$ and $\bar{\alpha}_A$ are defined by

$$\gamma^C_{AB} = \epsilon_{ABD}\eta^{DC} + 2\alpha_{[A}\delta^C_{B]}, \tag{2.11}$$

$$\bar{\eta}^{AB} \equiv \frac{1}{V}\int_{\mathcal{U}} \eta^{AB}\,dV, \tag{2.12a}$$

$$\bar{\alpha}_A \equiv \frac{1}{V}\int_{\mathcal{U}} \alpha_A\,dV; \tag{2.12b}$$

(c) a boundary condition

$$\frac{1}{V}\int_{\partial\mathcal{U}} \Delta^C_{AB}\epsilon_{BCM}\delta\theta^M\mathbf{e}^A\cdot d\mathbf{S} = 0 \tag{2.13}$$

for arbitrary $\delta\theta^M$, where $\delta\theta^M$ represents a small rotation of the $\{\mathbf{e}_A\}$ through an angle $(\delta\theta^M\delta\theta_M)^{1/2}$ about the axis $\mathbf{e}_M\delta\theta^M$, and $\partial\mathcal{U}$ is the boundary of $\mathcal{U}$.

Given a metric, we solve (2.7) for the vectors $\{\mathbf{e}_A\}$, the "best-fit" vectors, as functions of the spatial coordinates and the unknown C^C_{AB}. Then $\bar{\eta}^{AB}$ and $\bar{\alpha}_A$, which will have a parametric dependence on the C^C_{AB} or, equivalently, on the n^{AB} and a_A, are calculated and equations (2.9) and (2.10) are solved for n^{AB} and a_A. The simply-transitive group to which these structure constants correspond is said to be the approximate symmetry group of the metric. This group may be classified using the Bianchi scheme.[2] The metric is said to belong to a generalized Bianchi type (GBT).

Once the approximate symmetry group of g is known, we may construct a homogeneous approximant to g, that is, a homogeneous metric g_{homo} whose symmetry group is the same as the approximate symmetry group of g. We choose to do this as follows. Let $\{\mathbf{e}_A\}$ be the best-fit vectors and C^C_{AB} be the corresponding structure constants for g, found via the approximate symmetry technique. Also, let $\{\mathbf{E}_A\}$ be a set of invariant (under the approximate symmetry group) vector fields in $\mathcal{U}$ such that

$$[\mathbf{E}_A, \mathbf{E}_B] = C^C_{AB}E_C. \tag{2.14}$$

Once one of the many possible sets of $\{\mathbf{E}_A\}$ is chosen, we may define $(g_{\text{homo}})_{ab}$ as the matrix inverse of

$$g^{ab}_{\text{homo}} \equiv \delta^{AB}E^a_A E^b_B. \tag{2.15}$$

The vectors $\{\mathbf{E}_A\}$ will then be orthonormal basis vectors for g_{homo} and will generate the group reciprocal to the symmetry group of g_{homo}.

To fix the $\{\mathbf{E}_A\}$, we require that they differ as little as possible from $\{\mathbf{e}_A\}$, that is, that they minimize

$$K \equiv \frac{1}{V}\int_{\mathcal{U}} g_{ab}(E^a_A - e^a_A)(E^b_B - e^b_B)\delta^{AB}dV. \tag{2.16}$$

While restricting ourselves to $\{\mathbf{E}_A\}$ that satisfy (2.14), we vary[6] K with respect to the $\{\mathbf{E}_A\}$ and set the first variation equal to zero. This yields the differential equation

$$\nabla\cdot\{\mathbf{E}_B[(\mathbf{E}^B - \mathbf{e}^B)\cdot\mathbf{E}_A]\} + C^C_{AB}\mathbf{E}_C\cdot(\mathbf{E}^B - \mathbf{e}^B) = 0, \tag{2.17}$$

which must be satisfied by $\{\mathbf{E}_A\}$.

III. TWO-DIMENSIONAL SPACES

Two-dimensional spaces provide the simplest examples of our technique, for in two dimensions many computational difficulties disappear, along with, unfortunately, some of the technique's interesting aspects. Nonetheless, we include these examples because it is possible to give criteria permitting easy classification of all two-dimensional spaces, which is done in the two theorems below. Additionally, the two-dimensional spaces nicely illustrate some difficulties that can arise when the technique is applied globally, that is, when $\mathcal{U}=\mathcal{S}$.

In two dimensions ($N=2$), there are only two vectors in $\{\mathbf{e}_A\}$ and therefore only one commutator:

$$[\mathbf{e}_1, \mathbf{e}_2] = \gamma^1_{12}\mathbf{e}_1 + \gamma^2_{12}\mathbf{e}_2. \tag{3.1}$$

Only two types of groups are possible; their Lie algebras have the canonical forms[7]

$$[\mathbf{E}_1, \mathbf{E}_2] = 0 \tag{3.2}$$

and

$$[\mathbf{E}_1, \mathbf{E}_2] = C^1_{12}\mathbf{E}_1. \tag{3.3}$$

(When both structure constants are nonzero, a constant rotation transforms the commutator into the second form above). Further, there are no Jacobi conditions in two dimensions; any two constants can be the structure constants of a group. There is, therefore, no need for Lagrange multipliers.

Since all two-dimensional metrics are conformally flat,[8] we may take the metric to be

$$ds^2 = A^{-2}(x,y)(dx^2 + dy^2). \tag{3.4}$$

In terms of the Euler angle θ, we choose as orthonormal vectors

$$\mathbf{e}_1 = (A\cos\theta, A\sin\theta) \tag{3.5a}$$

$$\mathbf{e}_2 = (-A\sin\theta, A\cos\theta), \tag{3.5b}$$

whose commutation coefficients are

$$\gamma^A_{12} = \nabla\cdot\mathbf{e}_B\epsilon^{AB3}$$
$$= -\mathbf{e}_A\cdot(\nabla\theta + N), \tag{3.6}$$

where

$$N \equiv \left(A\frac{\partial A}{\partial y}, -A\frac{\partial A}{\partial x}\right). \tag{3.7}$$

Substituting (3.5) and (3.6) into (2.2) and performing the variations yields the two-dimensional versions of the differential and algebraic equations:

$$\nabla^2\theta = -\nabla\cdot\mathbf{N}, \tag{3.8}$$

or equivalently

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\theta = 0, \tag{3.9}$$

and

$$C^A_{12} = \frac{1}{V}\int_U \gamma^A_{12}\,dV, \tag{3.10}$$

together with the boundary condition

$$\frac{1}{V}\int_{\partial U}\delta\theta[\nabla\theta + \mathbf{N} + C^1_{12}\mathbf{e}_1 + C^2_{12}\mathbf{e}_2]\cdot d\mathbf{S} = 0. \tag{3.11}$$

That the structure constants are just averages of the commutation coefficients was to be expected from the absence of Lagrange multipliers. More surprising is the absence, in (3.9), of any dependence on the C^A_{12} or the metric. This is the principal degeneracy in two dimensions.

We wish to make our results here globally valid, so in what follows we shall set $U = S$ or, when appropriate, take a limit as $U \to S$ in some suitable fashion.

We are now able to obtain two very strong results, due to the simple form of γ^A_{12}.

Theorem 1: In a noncompact, two-dimensional, Riemannian manifold N, if the geometry is such that

$$\lim_{U\to N}\left[\frac{\text{surface area of } \partial U}{\text{volume of } U}\right] = 0, \tag{3.12}$$

where the limit is taken such that ∂U is always convex, then $C^A_{12} = 0$.

Proof: For noncompact spaces we use

$$C^A_{12} = \lim_{U\to N}\frac{1}{V}\int_U \gamma^A_{12}\,dV$$

which, after using (3.8), becomes

$$C^A_{12} = \lim_{U\to N}\frac{\epsilon^{AB3}}{V}\int_U \nabla\cdot\mathbf{e}_B\,dV$$
$$= \lim_{U\to N}\frac{\epsilon^{AB3}}{V}\int_{\partial U}\mathbf{e}_B\cdot\mathbf{n}\,dS.$$

Now $\mathbf{e}_B$ and $\mathbf{n}$ (the unit normal to ∂U) are unit vectors and must then satisfy

$$|\mathbf{e}_B\cdot\mathbf{n}| \leq 1.$$

This implies that

$$[\text{surface area of } \partial U] = \int_{\partial U} dS \geq \left|\int_{\partial U}\mathbf{e}_B\cdot\mathbf{n}\,dS\right|$$

and therefore

$$\frac{\text{surface area of } \partial U}{\text{volume of } U} = \frac{1}{V}\int_{\partial U} dS \geq \frac{1}{V}\left|\int_U \gamma^A_{12}\,dV\right|;$$

hence the result stated. QED

Convincing oneself that there exist two-dimensional spaces in which the ratio in Theorem 1 has a nonzero limit is not difficult. One example is the Poincaré half-plane,[9] which is a homogeneous space of constant negative curvature. Small perturbations of this space must yield inhomogeneous spaces which also have a nonzero limit for (3.12).

An even stronger result is possible when considering compact spaces:

Theorem 2: If the manifold is compact and orientable, then $C^A_{12} = 0$.

Proof: Using Green's theorem,[10]

$$C^A_{12} = \frac{1}{V}\int_N \gamma^A_{12}\,dV$$
$$= \frac{\epsilon^{AB3}}{V}\int_N \nabla\cdot\mathbf{e}_B\,dV$$

since the boundary of a boundary is zero. QED
Theorem 2 shows that, with this scheme, all compact orientable 2-spaces have the same approximate symmetry group, regardless of their geometry; the topology determines the classification.

What are the homogeneous approximants for two-dimensional spaces? If the approximant is required to be orientable and have a simply transitive, two-parameter group defined globally, there are only four admissible approximants[11] (aside from scale changes): (i) the noncompact space of constant negative curvature, (ii) the flat Euclidean plane, (iii) the cylinder, and (iv) the flat 2-torus. The other possible spaces are either not orientable or have no globally defined two-parameter group. For example, the 2-sphere with constant positive curvature is invariant under a multiply transitive, three-parameter group which has no simply transitive, two-parameter subgroups. Often, then, it will be impossible to find a global approximant for an inhomogeneous two-dimensional space.

Though mathematically simple, the two-dimensional version of our technique is really too simple to yield much insight into the method. Also, the nonparallelizability of some compact 2-spaces gives rise to problems which do not occur when dealing with 3-spaces. We therefore turn to three-dimensional spaces.

IV. THE GOWDY T^3 METRIC: A SPACETIME EXAMPLE

The Gowdy T^3 spacetimes[12,13] are solutions to the vacuum Einstein equations with inhomogeneous spacelike sections that are topologically 3-tori. These spacetimes contain no preferred family of spacelike hypersurfaces. Hence, the choice of hypersurfaces to

be analyzed by the approximate symmetry technique is somewhat arbitrary. We investigate here two possible families and, within each family, study the dependence of I and n^{AB} and a_A as the hypersurfaces evolve. The inhomogeneity in these spacetimes arises from the presence of gravitational waves, which circumnavigate the torus in one direction; members of this family of solutions differ in the waveforms they contain. Although inhomogeneous, they possess some symmetry; a two-parameter, Abelian group acts in the (spacelike) planes perpendicular to the direction of wave propagation. In addition to the compactness of the spacelike sections and the presence of a two-parameter symmetry group, one other property makes them convenient for our purposes: They may be covered by a single set of coordinates. (In a strict sense, more than one coordinate patch is needed, because of identifications made in the covering manifold). This permits us to avoid many inessential—but troublesome—coordinate problems and devote our efforts to applying the approximate symmetry technique.

The spacetime metric is

$$ds^2_{(4)} = A^2(t,\theta)[-dt^2 + d\theta^2] + 2t[D^2(t,\theta)d\sigma^2 + D^{-2}(t,\theta)d\delta^2], \tag{4.1}$$

where $0 \leq \theta \leq 2\pi$ and $\oint\!\!\oint d\sigma d\delta = 16\pi$. The metric coefficients are periodic in θ with a period of 2π, permitting identification of points at $(\theta,\sigma,\delta) = (0, \sigma, \delta)$ and $(2\pi,\sigma,\delta)$. Identifications in σ and δ are made in a similar way. The absence of any metric dependence upon σ and δ is due to the two-parameter symmetry group, whose orbits are the tori given by: $t = \text{const}$, $\theta = \text{const}'$. Following Berger,[14] we may write

$$A = (2t)^{-1/4}\exp(-3B_+(t,\theta)/2,$$
$$D = \exp(\sqrt{3}\,B_-(t,\theta). \tag{4.2}$$

The function B_- can be expressed as a sum of standing gravitational waves. The resulting expression for B_- is given in Appendix B, wherein we summarize many of the detailed expressions concerning this metric. The nonlinearities in the solution have been lumped into the other metric function, B_+, which depends, in part, upon sums and differences of the modes in B_-. Inspection of the limiting forms for the metric coefficients in Appendix B shows that, when considered as evolving 3-spaces, the Gowdy models have a singular (zero volume) beginning at $t = 0$ and expand indefinitely thereafter.

We will focus attention on two families of spacelike hypersurfaces. The first family, $\mathcal{S}_G(t)$, will be that picked out by the coordinates in which the metric is expressed, that is, the family given by $t = \text{const}$. These hypersurfaces allow analytical calculation of most of the quantities of interest. The normals to $\mathcal{S}_G$ are not, however, geodesic. The acceleration of the unit normal

$$e_0^\alpha = A^{-1}\delta_0^\alpha \tag{4.3}$$

[in the coordinates of (4.1)] is found to be

$$\dot{e}_0^\alpha \equiv e_0^\mu \nabla_\mu e_0^\alpha = A^{-3}\frac{\partial A}{\partial \theta}\,\delta_1^\alpha, \tag{4.4}$$

which is in general a nonzero vector in the θ direction. The acceleration is zero only: (i) at isolated values of θ, (ii) in the limit as $t \to \infty$, and (iii) in the limit as $t \to 0$ for those metrics which satisfy $1 - 3a^2 > 0$, where $a(\theta)$ is defined in (B7). Because of their obvious geometrical significance, families of geodesic-normal hypersurfaces are more natural than an arbitrarily chosen family such as $\mathcal{S}_G(t)$. Therefore, we shall also want to use a geodesic-normal family $\mathcal{S}'_G(J)$, when one exists,[15] for some calculations. The details of how this family is obtained have been put in Appendix C; here we will just give enough of these calculations so that the results, summarized in Table I, can be under understood.

At any given space—time point, the normal $\mathbf{e}'_0$ to the member of $\mathcal{S}'_G$ passing through that point will be related to $\mathbf{e}_0$ at that point by a Lorentz transformation in the t—θ plane. Thus we may express $\mathbf{e}'_0$ as

$$\mathbf{e}'_0 = (\cosh\psi, \sinh\psi, 0, 0)/A, \tag{4.5}$$

where the boost angle ψ is a function of t and θ. The equation for the hypersurface $t = f(\theta, J)$, is then determined through

TABLE I. Summary of the properties of $\mathcal{S}_G$ and $\mathcal{S}'_G$.

	Coordinate-determined hypersurfaces $\mathcal{S}_G$	Geodesic-normal hypersurfaces $\mathcal{S}'_G$
Normal vectors	$\mathbf{e}_0 = (1, 0, 0, 0)/A$	$\mathbf{e}'_0 = (\cosh\psi, \sinh\psi, 0, 0)/A$, where $\psi = \psi(t, \theta)$ satisfies $\frac{\partial}{\partial t}(A\sinh\psi) = -\frac{\partial}{\partial\theta}(A\cosh\psi)$
Equation of hypersurfaces	$t = \text{const}$ t parametrizes family	$t = f(\theta, J)$, where $\partial f/\partial\theta = \tan\psi(f(\theta, J), \theta)$ J parametrizes family
Are the hypersurfaces geodesic-normal?	Only in the limit as $t \to \infty$ and, if $1 - 3a^2(\theta) > 0$, in the limit $t \to 0$.	Yes, always. They are chosen to coincide with $\mathcal{S}_G(t)$ in the limit $t \to \infty$ and, if $1 - 3a^2 > 0$, in the limit $t \to 0$.
Metric on the hypersurfaces	$ds^2 = A^2(t, \theta)\,d\theta^2 + 2t[D^2(t, \theta)\,d\sigma^2 + D^{-2}(t, \theta)\,d\delta^2]$	$ds' = A^2(f, \theta)[1 - (\partial f/\partial\theta)^2]\,d\theta^2 + 2f[D^2(f, \theta)\,d\sigma^2 + D^{-2}(f, \theta)\,d\delta^2]$
Differential, proper-volume scalar	$dV = 2tA\,d\theta\,d\sigma\,d\delta$	$dV' = 2fA(f, \theta)[1 - (\partial f/\partial\theta)^2]^{1/2}\,d\theta\,d\sigma\,d\delta$ $= [2tA(f, \theta)/\cosh\psi(f, \theta)]\,d\theta\,d\sigma\,d\delta$

$$\frac{\partial f}{\partial \theta} = \tanh\psi(f,\theta). \tag{4.6}$$

We ensure that $\mathbf{e}_0'$ is tangent to geodesic curves by choosing ψ to satisfy the geodesic equation:

$$\frac{\partial}{\partial t}[A\sinh\psi] + \frac{\partial}{\partial\theta}[A\cosh\psi] = 0. \tag{4.7}$$

For definiteness, we henceforth consider only those solutions to (4.7) which have the asymptotic behavior $\psi \to 0$ as $t \to \infty$, so $\mathbf{e}_0$ and $\mathbf{e}_0'$ (and, therefore, S_G and S_G') coincide in that limit. For those metrics which satisfy $1 - 3a^2(\theta) > 0$, we will require, in addition, that $\psi \to 0$ as $t \to 0$; thus $\mathbf{e}_0$ and $\mathbf{e}_0'$ coincide in that limit also.

We shall begin by applying the approximate symmetry technique to the hypersurfaces S_G. Rather than solve the differential equation by systematic integration, we simply guess a form for the triad and see whether it solves the equations.

On the S_G hypersurfaces, the obvious triad of orthonormal vectors to try is one which incorporates the extant symmetry group; thus we pick

$$e_1^a = A^{-1}\delta_1^a, \quad e_2^a = (2tD)^{-1}\delta_2^a, \quad e_3^a = (D/2t)\delta_3^a. \tag{4.8}$$

The only nonzero commutation coefficient is

$$\eta^{23} = -(AD)^{-1}\frac{\partial D}{\partial\theta} = -\sqrt{3}\,(2t)^{1/4}\frac{\partial B_*}{\partial\theta}\exp(3B_*/2). \tag{4.9}$$

For this triad we obviously have $\bar\eta\cdot\bar\alpha = 0$; thus the algebraic equations must yield $n^{23} = \bar\eta^{23}$, $n^{AB} = 0 = a_A$ otherwise. In fact, even n^{23} is zero, since

$$\eta^{23} = -\frac{32\pi t}{V}\int_0^{2\pi}\frac{\partial}{\partial\theta}(\ln D)\,d\theta = 0 \tag{4.10}$$

due to the periodicity of the metric. The differential equation, which now reduces to

$$\nabla\cdot(\mathbf{e}^A\gamma^C_{AB})\epsilon_{BCM} = 0, \tag{4.11}$$

is identically satisfied, as is the boundary condition, (2.13), since the hypersurfaces are compact. Therefore (4.8), together with the structure constants $C^C_{AB} = 0$, gives an extremum of the variational integral.[16] We conclude that the approximate symmetry group of these hypersurfaces is of GBT I.

Does (4.8) yield a minimum for I, rather than a maximum or inflection point? The question is a knotty one due to the complexity of the expression for the second variation of I. No decisive general answer has emerged, despite a number of attacks on the problem. The restricted problem, in which only those Gowdy metrics that are almost homogeneous (i.e., the wave amplitudes are small) are considered, admits an easy solution, though. In this case, results[1] concerning the linearized variational problem may be applied; we find that, indeed, (4.8) gives a minimum.

We now turn to the temporal behavior of I at its minimum. In terms of (η, α) and (n, α), the value of I at a minimum is

$$\mathrm{I} = \frac{1}{V}\int 2(\eta^2 + 2\alpha^2)\,dV - 2(n^2 + 2a^2), \tag{4.12}$$

which reduces to

$$\mathrm{I} = 16\pi\frac{8t}{V}\int_0^{2\pi}\frac{1}{A}\left(\frac{\partial}{\partial\theta}\ln D\right)^2 d\theta. \tag{4.13}$$

It is revealing to examine the behavior of I in both the asymptotic future and near the initial singularity. The asymptotic expressions (B9) and (B10) for the metric coefficients show that A and D are independent of θ asymptotically; therefore,

$$\lim_{t\to\infty}\mathrm{I} = 0. \tag{4.14}$$

Intrinsically, the $S_G(t)$ become more homogeneous as t gets large. This accords with our intuition about this metric; the hypersurface volume monotonically increases with time, so the energy density of the gravitational waves, which (asymptotically) behaves[17] as

$$[\text{energy density}] \sim V^{-4/3}, \tag{4.15}$$

decreases with time; hence the inhomogeneity-producing waves become less effective and the metric more homogeneous.

The behavior near the singularity is surprising. When the limiting expressions (B6)—(B8) for this regime are substituted into (4.13), we get

$$\mathrm{I} \sim 12(\ln 2t)^2\,\frac{\int_0^{2\pi} d\theta\,(da/d\theta)^2(2t)^{(1-3a^2)/4}}{\int_0^{2\pi} d\theta\,(2t)^{-(1-3a^2)/4}}, \tag{4.16}$$

so

$$\lim_{t\to 0}\mathrm{I} = \begin{cases} \infty & \text{if } 1 - 3a^2 < 0, \\ 0 & \text{if } 1 - 3a^2 > 0. \end{cases} \tag{4.17}$$

The S_G for those metrics having $1 - 3a^2 > 0$ become intrinsically homogeneous as the singularity is approached.[18] (This is the same criterion that must be satisfied if S_G is to be geodesic-normal in that limit). The condition is sufficient but not necessary. It was obtained assuming the metric coefficients contained both Bessel and Neumann functions. If only Bessel functions appear, $\lim_{t\to 0}\mathrm{I} = 0$ regardless of the sign of $1 - 3a^2$. Using a computer, we have calculated I(t) for three sample metrics with two wave modes excited. The results are given in Fig. 1.

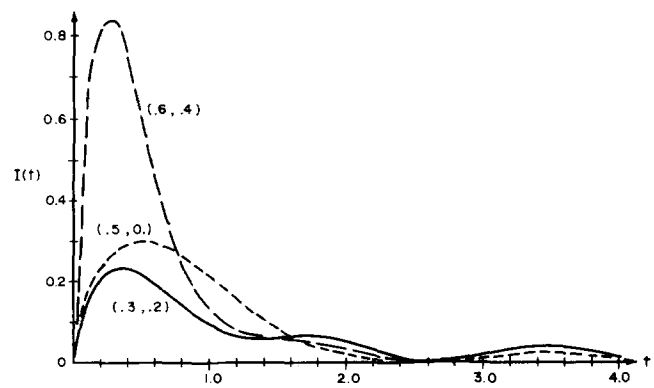


FIG. 1. The values of I(t), when the gravitational wave modes $n = 1$ and $m = 2$ have been excited, are plotted. For all three cases we have assumed that Z_1 contains no Neumann function and $\theta_1 = 0°$ and $\theta_2 = 40°$. The numbers in parentheses are (c_1, c_2).

How is this strange behavior explained? The hypersurfaces in S_G contract to zero volume as $t \to 0$; Thus an explanation similar to that given for the limit as $t \to \infty$ seems to be impossible. Indeed, that argument in reverse seems to imply that I must become infinite as $t \to 0$. On closer examination, we find that the behavior as $t \to 0$ has an explanation similar to that for $t \to \infty$. Note the "shape" of the singularity; in all cases a small spherical volume is stretched into a cigar at t goes to zero. If $1 - 3a^2 < 0$, the cigar's long axis is in the δ direction; if $1 - 3a^2 > 0$, the long axis points along the θ direction. Recalling that the gravitational waves travel in the θ direction only, we see that stretching in this direction would tend to wash out the inhomogeneous ripples in the geometry, thus making the space more homogeneous.

The S_G have been found to be GBT I, but is that typical of families of hypersurfaces in the Gowdy metrics? What would the approximate symmetry technique yield if another family of hypersurfaces were used? To partially answer that question, we turn to the second family mentioned, S_G'.

Again we pick the simplest triad incorporating the symmetry group and find that it solves the differential equation and boundary condition. That triad $\{\mathbf{e}_A'\}$ is:

$$\mathbf{e}_1' = (\sinh\psi, \cosh\psi, 0, 0,)/A, \tag{4.18a}$$

$$\mathbf{e}_2' = (0, 0, 1, 0)/2tD, \tag{4.18b}$$

$$\mathbf{e}_3' = (0, 0, 0, 1)D/2t. \tag{4.18c}$$

The nonzero commutation coefficients are

$$\eta'^{23} = -\left(\frac{\partial D}{\partial t}\sinh\psi + \frac{\partial D}{\partial \theta}\cosh\psi\right)\Big/AD, \tag{4.19}$$

$$\alpha'^{1} = -\frac{1}{2tA}\sinh\psi. \tag{4.20}$$

On each hypersurface, $t = f(\theta, J)$,

$$\eta'^{23} = \eta'^{23}(f(\theta, J), \theta) \tag{4.21}$$

and similarly for α'^{1}. Using (4.6), we may rewrite (4.19) as

$$\eta'^{23} = -\frac{1}{AD}\frac{dD}{d\theta}(f(\theta, J), \theta)\cosh\psi. \tag{4.22}$$

The structure constants are again simply equal to the averages of η'^{AB} and α_A'. We find that they are all zero[19] except for n'^{23}, which is

$$n'^{23} = \frac{2}{(V'/16\pi)}\int_0^{2\pi} d\theta \ln D(f, \theta)\tanh\psi(f, \theta). \tag{4.23}$$

To evaluate (4.23), we must know the dependence of f upon θ; this entails integration of the geodesic equation. As noted in Appendix C, a complete integral for the geodesic equation cannot be found in general. Consequently, further restrictions are required before detailed knowledge of n'^{23} can be obtained. For these metrics satisfying $1 - 3a^2 > 0$ and containing weak gravitational waves, S_G and S_G' will differ only slightly and will coincide in the limits as $t \to 0$ and ∞. A restriction to such metrics would allow linearization of the geodesic equation for ψ; the exact solution for ψ could be approximated by the solution of the linearized equation. Henceforth, we will limit our considerations to such metrics and do all calculations to first order in ψ. With these provisos, the previous statements concerning I and n^{23} for S_G in the limits as $t \to 0$ and ∞ hold qualitatively for S_G' also.

Under the above assumptions, the solutions for ψ and f are (from Appendix C):

$$\psi(f, \theta) \approx -A^{-1}(t, \theta)\int_0^t dt' \frac{\partial A}{\partial \theta}(t', \theta), \tag{4.24}$$

and

$$f(t, \theta) = J + \int_0^\theta d\theta\, \psi(t, \theta) + O(\psi^2). \tag{4.25}$$

Even with (4.24) and (4.25) in hand, we find the integral in (4.23) still cannot be done analytically, except for those metrics with only one gravitational wave mode excited. In that case, the integrand can be expressed as perfect differential, whose integral vanishes due to the periodicity of the metric coefficients: again we get GBT I.

Nevertheless, we expect n'^{23} to be nonzero in general, giving GBT VI_0; for to first order in ψ, the parameter n'^{23} essentially gives the inner product between the "vectors" B_- and ψ in Fourier transform space. Using $\ln D = \sqrt{3}B_-$, (B4), and the Fourier expansion of ψ, i.e.,

$$\psi = \tfrac{1}{2}\varphi_0 + \sum_{m=1}^{\infty}\varphi_m(t)\cos[m(\theta - \theta_m)], \tag{4.26}$$

(4.23) becomes

$$n'^{23} = \frac{2\sqrt{3}\pi}{(V'/16\pi)}\sum_{m=1}^{\infty} c_n Z_0(mt)\varphi_m(t) + O(\psi^2). \tag{4.27}$$

Hence, ψ and B_- must be orthogonal for n'^{23} to be zero, a situation unlikely in general.

Using (4.24) and (4.25), we have evaluated (4.23) numerically (to first order in ψ) when one and two modes are excited. The results, shown in Figs. 2 and 3, indicate that n'^{23} oscillates between positive and negative values, the oscillations damping out in the asymptotic future.

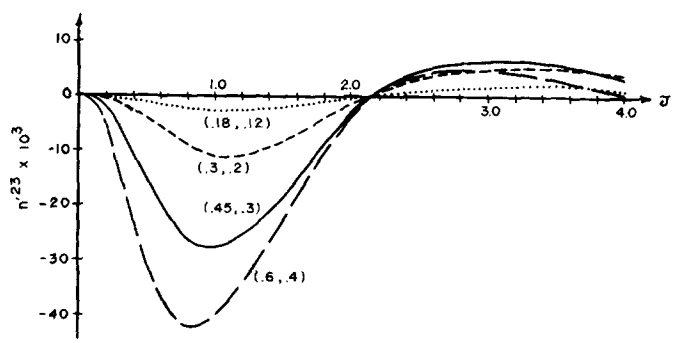


FIG. 2. Plotted are the results of a computer calculation of n'^{23} to first order in ψ for various wave amplitudes. We have assumed that only the modes $n=1$ and $n=2$ are excited and that the coefficients of the Neumann functions are zero. The values of θ_i are $\theta_1 = 0°$ and $\theta_2 = 40°$. The numbers in parentheses give (c_1, c_2). The case where either $c_1 = 0$ or $c_2 = 0$ yields $n'^{23} = 0$. Note the earlier peaking of plots with larger values of c_1 and c_2.

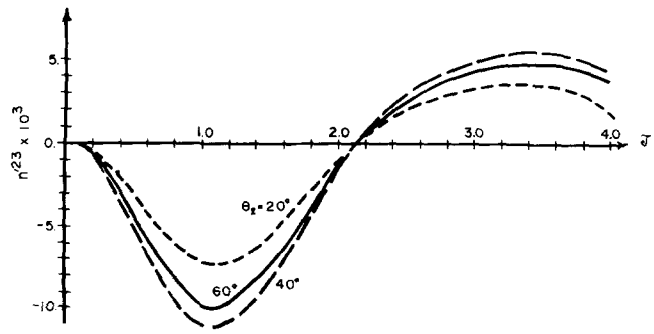


FIG. 3. Here we have made the same assumptions as in the previous figure but have varied θ_2 instead of c_1 and c_2. The values chosen for the nonzero c_n are $c_1 = 0.3$ and $c_2 = 0.2$. Cases with $\theta_1 = 0°$ and 90° given $n'^{23} = 0$ to within the accuracy of the calculations.

We return to the question of the generic classification of spacelike hypersurfaces in the Gowdy metric. Note that the triad $\{\mathbf{e}'_A\}$ in (4.18) solves the differential equation for arbitrary ψ; we did not have to require that ψ solve the geodesic equation. The crucial condition was that ψ (and, hence, the equation for the hypersurface, $t = f$) was independent of σ and δ. Therefore, we may say that, for any spacelike hypersurface whose metric is independent of σ and δ, the triad (4.18) solves (2.7) and (2.13), and therefore the hypersurface can be classified GBT I or VI_0, depending upon whether n'^{23}, as given in (4.23) is zero or nonzero.

As an addendum to studying the approximate symmetries of the intrinsic geometry of the hypersurfaces in $\mathcal{S}_G$ and $\mathcal{S}'_G$, we wish to turn our attention briefly to the extrinsic geometry of these hypersurfaces. We ask the question: To what extent is the extrinsic geometry invariant under the approximate symmetry group? Here we will restrict our consideration to the limits where $t \to 0$ or ∞.

The extrinsic curvature of a hypersurface tells one how the hypersurface is embedded in the space–time and can be represented as a tensor defined on the hypersurface itself. For a family of hypersurfaces, it specifies the differential change in a spherical volume from one hypersurface to the next. The triad components of the extrinsic curvature tensor are given by[2]

$$\Theta_{AB} = -\mathbf{e}_B \cdot [\mathbf{e}_0, \mathbf{e}_A], \tag{4.28}$$

where $\mathbf{e}_0$ is the unit normal to the hypersurface. For $\mathcal{S}_G$, this is explicitly

$$\Theta_{AB} = \mathrm{diag}\left(A^{-2}\frac{\partial A}{\partial t}, A^{-1}\frac{\partial}{\partial t}[\ln\sqrt{2t}D], -A^{-1}\frac{\partial}{\partial t}[\ln(D/\sqrt{2t})]\right). \tag{4.29}$$

If the space–time were spatially homogeneous and the hypersurfaces were the group orbits, Θ_{AB} would be a function of time alone and

$$\mathcal{L}_{\mathbf{e}_L}\Theta_{AB} = 0. \tag{4.30}$$

Thus we might require that the normalized derivative satisfy[20]

$$\left|\frac{\mathcal{L}_{\mathbf{e}_L}\Theta_{AB}}{\Theta_{AB}}\right| \leq I \tag{4.31}$$

if Θ_{AB} is to be approximately invariant to the same extent as g. In the limit, we will want

$$\lim_{t\to\infty}\left[\frac{\mathcal{L}_{\mathbf{e}_L}\Theta_{AB}}{\Theta_{AB}}\right] = 0 \tag{4.32}$$

since I becomes zero in this limit. If (4.32) obtains, we may say that the space–time is spatially homogeneous in this limit. A similar requirement is made in the limit as $t \to 0$ for those metrics having $\lim_{t\to 0} I = 0$. Using the asymptotic expressions we find, indeed, that (4.32) is satisfied. Thus we may say that the Gowdy space–times become spatially homogeneous in the asymptotic future. (This is true for both families of hypersurfaces previously considered, since $\mathcal{S}'_G$ coincides with $\mathcal{S}_G$ in this limit).

Near the singularity[21]

$$\lim_{t\to 0}\frac{\mathcal{L}_{\mathbf{e}_L}\Theta_{AB}}{\Theta_{AB}} = 0, \tag{4.33}$$

for those metrics satisfying $1 - 3a^2 > 0$. Therefore, these metrics become spatially homogeneous near the singularity. For the metrics with $1 - 3a^2 < 0$,

$$\lim_{t\to 0}\left|\frac{\mathcal{L}_{\mathbf{e}_L}\Theta_{AB}}{\Theta_{AB}}\right| = \infty, \tag{4.34}$$

again mimicking the behavior of the intrinsic geometry. As before, the behavior of the two families $\mathcal{S}_G$ and $\mathcal{S}'_G$ is the same when $1 - 3a^2 > 0$ obtains. When $1 - 3a^2 < 0$, it is likely that the family $\mathcal{S}'_G$ cannot be defined arbitrarily close to the singularity. To avoid this (inessential) difficulty, we shall not consider $\mathcal{S}'_G$ in this limit when $1 - 3a^2 < 0$. Results concerning $\mathcal{S}_G$ and $\mathcal{S}'_G$ are summarized in Table II.

V. HOMOGENEOUS APPROXIMANTS FOR $\mathcal{S}_G$ and $\mathcal{S}'_G$

We begin with the homogeneous approximant for $\mathcal{S}_G$.

The hypersurfaces in $\mathcal{S}_G$ are GBT I; so we want to form an approximant whose symmetry group has the Lie algebra

$$[\mathbf{E}_A, \mathbf{E}_B] = 0. \tag{5.1}$$

TABLE II. The extent to which the intrinsic and extrinsic geometries of $\mathcal{S}_G$ and $\mathcal{S}'_G$ are invariant under the approximate symmetry groups is indicated here. In those cases marked by an asterisk, the surfaces $\mathcal{S}'_G$ are not considered

	Metric satisfies: $1 - 3a^2 > 0$	Metric satisfies: $1 - 3a^2 < 0$
Bianchi type of the approximate symmetry group	I for $\mathcal{S}_G$; VI_0 for $\mathcal{S}'_G$ (locally)	same as for $1 - 3a^2 > 0$
$\lim I$ $t \to 0$	zero	infinite*
$t \to \infty$	zero	zero
$\lim \lvert\mathcal{L}\Theta\rvert/\Theta$ $t \to 0$	zero	infinite*
$t \to \infty$		zero
singularity	cigar along θ	cigar along δ

In terms of the matrix inverses E^A_a, this equation is

$$\epsilon^{abc}\frac{\partial}{\partial x^b}E^A_c = 0 \tag{5.2}$$

and has the solution

$$E^A_a = \frac{\partial u^A}{\partial x^a}, \tag{5.3}$$

where u^A are arbitrary functions of the coordinates $(x^a) = (\theta, \sigma, \delta)$. Making the reasonable assumption that the vectors $\{\mathbf{E}_A\}$ lie along the coordinate lines, so that E^A_a is diagonal, gives

$$\frac{\partial u^A}{\partial x^a} \propto \delta^A_a. \tag{5.4}$$

We conclude that E^a_A is of the form

$$E^a_A = \mathcal{E}_A(x^A)\delta^a_A \quad \text{(no sum)}. \tag{5.5}$$

The differential equation (2.17) connecting $\{\mathbf{E}_A\}$ with $\{\mathbf{e}_A\}$ becomes

$$\frac{\partial}{\partial x^B}[g_{BB}\sqrt{g}\,(\mathcal{E}_B)^2(\mathcal{E}_B - e^B_B)] = 0 \quad \text{(no sum)}. \tag{5.6}$$

For $B=2$, this reduces further to

$$g_{22}\sqrt{g}\,(3\mathcal{E}_2 - 2e^2_2)\mathcal{E}_2\frac{\partial \mathcal{E}_2}{\partial \sigma} = 0, \tag{5.7}$$

requiring

$$\mathcal{E}_2 = \text{const} \tag{5.8}$$

Similarly, letting $B=3$ in (5.6) gives

$$\mathcal{E}_3 = \text{const}' \tag{5.9}$$

The integral of Eq. (5.6) for $B=1$ is

$$(\mathcal{E}_1)^2(\mathcal{E}_1 - e^1_1) = \text{const}''/g_{11}\sqrt{g}\,. \tag{5.10}$$

This can have up to three real roots, depending upon const'' and $g_{11}\sqrt{g}$. The value of the constants in (5.8), (5.9), and (5.10) are fixed by returning to the integral K which was varied and requiring $\delta K = 0$. We find

$$\mathcal{E}_1 = e^1_1 = 1/A \tag{5.11}$$

(i.e., const'' $=0$),

$$\mathcal{E}_2 = \frac{\int dV\, g_{22}e^2_2}{\int dV\, g_{22}} = \frac{\int dV\,\sqrt{2t}\,D}{\int dV\, 2tD^2}, \tag{5.12}$$

and

$$\mathcal{E}_3 = \frac{\int dV\, g_{33}e^3_3}{\int dV\, g_{33}} = \frac{\int dV\,\sqrt{2t}/D}{\int dV\, 2t/D^2}. \tag{5.13}$$

The approximant is therefore

$$ds^2_{\text{homo}} = A^2d\theta^2 + (\mathcal{E}_2)^{-2}d\sigma^2 + (\mathcal{E}_3)^{-2}d\delta^2. \tag{5.14}$$

It is instructive to consider this expression in the limit as $t\to\infty$. Using the asymptotic expressions in Appendix B, one finds that (5.14) reduces to

$$ds^2_{\text{homo}_\infty} = (2t)^{-1/2}\exp(-3B_+)\,d\theta^2 + r^2(t)(2t)^{(1+\sqrt{3}a_0)}d\sigma^2 + r^2(t)(2t)(1-\sqrt{3}\,a_0)\,d\delta^2, \tag{5.15}$$

where

$$r(t) \equiv \frac{\int_0^{2\pi} d\theta\,\exp(2\sqrt{3}\,q(t,\theta))}{\int_0^{2\pi} d\theta\,\exp(\sqrt{3}\,q(t,\theta))}, \tag{5.16}$$

$$q(t,\theta) \equiv B_- - \tfrac{1}{2}a_0\ln 2t$$

$$= 2(\pi 2t)^{-1/2}\sum_{n=1}^{\infty}(n)^{-1/2}\cos[n(\theta-\theta_n)]. \tag{5.17}$$

If the factors of r^2 were set equal to 1, (5.15) would be identical to the spatial part of the (homogeneous) background metric Berger found for the Gowdy spacetime in this limit. Because of the r^2 in (5.15), our approximant is not a solution to the Einstein equations. This is not surprising: Nothing in our averaging scheme guarantees that a solution results.

We encounter difficulties when we try to construct an approximant for the S'_G hypersurfaces. The Lie algebra for that approximant is

$$[\mathbf{E}'^*_1, \mathbf{E}'^*_2] = n'^{23}\mathbf{E}'^*_2, \tag{5.18a}$$

$$[\mathbf{E}'^*_1, \mathbf{E}'^*_3] = -n'^{23}\mathbf{E}'^*_3, \tag{5.18b}$$

$$[\mathbf{E}'^*_1, \mathbf{E}'^*_3] = 0. \tag{5.18c}$$

These equations could be solved in much the same fashion as was done for (5.1); but we would find it impossible to require simultaneously that the solutions be periodic in θ and vanish nowhere. It seems that one cannot impose a 3-torus topology upon a homogeneous metric of BT VI_0. This is confirmed by Theorem 2.10 of Yano and Bochner,[10] which states that, in a compact Riemannian manifold, there exists no nonzero Killing vector field satisfying

$$\xi \circ \mathbf{R}^* \circ \xi \leq 0. \tag{5.19}$$

The triad components of the Ricci tensor for our approximant must be[22]

$$R^*_{AB} = \text{diag}(-2[n'^{23}]^2, 0, 0); \tag{5.20}$$

therefore, at least one member of a triplet of linearly independent Killing vector fields must obey (5.19). The conditions of Theorem 2.10 are satisfied: We must conclude that the type VI_0 approximate symmetry group can only be a local group; a global approximant is impossible.

VI. DISCUSSION

In our previous paper,[1] we noted that the magnitude of I at its minimum may be used as a measure of a given 3-metric's inhomogeneity: The greater I, the more inhomogeneous is g, and vice-versa. (one has $I=0$ if and only if the metric is homogeneous.) Roughly, I may be thought of as measuring the "perpendicular distance" in superspace from the 3-metric in question to the nearest submanifold of homogeneous metrics. Further, we can distinguish between two types of inhomogeneity: those in which $\lambda_A = 0$ and those with $\lambda_A \neq 0$.

In light of these remarks, two comments come to mind concerning the preceding examples. First, we have not yet found an inhomogeneous solution to the Einstein equations in which $\lambda_A \neq 0$. It is tempting to conjecture that the symmetries of the metrics studied preclude $\lambda_A \neq 0$, but whatever its explanation, this fact highlights the specialized nature of the two inhomogeneous solutions investigated.

The second comment relates to the asymptotic behavior of I in the Gowdy example. We found that,

as $t \to \infty$, $I \to 0$; that is, the intrinsic geometry of the hypersurfaces became more homogeneous as $t \to \infty$. One wonders whether this behavior is typical of cosmological solutions that do not collapse or whether it is peculiar to the Gowdy model. If the behavior is atypical, what characterizes those solutions displaying asymptotic intrinsic homogeneity?

In the Gowdy example, the extrinsic geometry also became more homogeneous as $t \to \infty$. We suspect that this is not characteristic of more general solutions, for there exists a class of Szekeres solutions[23] which are, intrinsically, exactly homogeneous but are extrinsically inhomogeneous.

Finally, we remark that one may wish to extend our scheme to include a slicing algorithm for spacetimes. For example, the choice of hypersurfaces may be incorporated into the variational principle by requiring that the slicing minimize the integral of I over the spacetime. In this way we may obtain a measure of the total inhomogeneity of the spacetime.

APPENDIX A: KANTOWSKI-SACHS SPACES

The Kantowski—Sachs spacetimes[4] admit a multiply-transitive, four-parameter Lie group acting on space-like hypersurfaces. The group has no three-parameter, simply-transitive subgroups; so, although the hypersurfaces are homogeneous in a physical sense, they do not satisfy the stricter definition given in Sec. II, and an invariant triad cannot be found. Therefore, it is of some interest to analyze these hypersurfaces with the appproximate symmetry technique: They could provide a useful testpiece for questions on the behavior of our scheme in the presence of metric symmetries.

In the coordinates of Kantowski and Sachs, the metric of the hypersurfaces on which the group acts is

$$ds^2 = X^2 d\rho^2 + Y^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{A1}$$

where X and Y may depend upon the hypersurface in question, but are independent of ρ, θ, and ϕ. The topology of these hypersurfaces is $R \times S^2$; the coordinate ρ, with $(-\infty < \rho < +\infty)$, labels the 2-spheres; θ and ϕ, with $(0 \leq \theta \leq \pi,\ 0 \leq \phi \leq 2\pi)$, label points on the 2-spheres. The metric resembles that of Euclidean 3-space in spherical-polar coordinates, but instead of scaling the geometry on each 2-sphere according to its distance from the origin, each 2-sphere is given the same geometry.

Cartesian coordinates will facilitate our analysis; we transform to them in two steps. The half-space consisting of 2-spheres with $\rho \geq 0$ may be mapped into R^3 from which points within a sphere of radius 1 about the origin have beem removed. If the spherical-polar coordinates (r, θ', ϕ') are used in R^3, the mapping is given by

$$r = \exp(X\rho/Y), \quad \theta' = \theta, \quad \phi' = \phi. \tag{A2}$$

The metric now becomes

$$ds^2 = (Y^2/r^2)[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{A3}$$

where $1 \leq r \leq \infty$ and the primes have been dropped. The spherical-polar coordinates may be related to Cartesian coordinates in the standard way to give

$$ds^2 = (Y^2/r^2)(dx^2 + dy^2 + dz^2). \tag{A4}$$

The other half-space, with $\rho \leq 0$, may be embedded in an identical R^3 in a similar way. The two R^3 sheets are then identified on their $r = 1$ surfaces to give the complete Kantowski—Sachs 3-spaces (see Fig. 4).

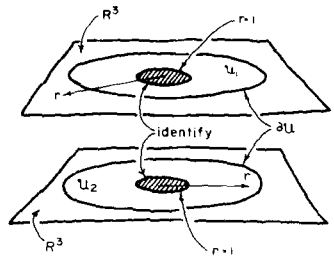


FIG. 4. The embedding of the Kantowski—Sachs space into two R^3 spaces is shown (with one dimension suppressed). The shaded areas have been deleted and the indicated identification made. The set U is given by $U = U_1 + U_2$. The metric on each sheet is (A4).

To solve the variational principle, we note that (A4) is spherically symmetric—there is no essential distinction between the three coordinate directions. This leads us to choose (in the x, y, and z coordinates)

$$e^a_A = (r/Y)\delta^a_A, \tag{A5}$$

as a possible triad of orthonormal vectors.

In practice, it proves useful to find those commutation coefficients whose averages may be nonzero, thus permitting at least a partial solution to the algebraic equations before one turns to the differential equations. For the triad of (4.5), we find

$$\gamma^C_{AB} = 2x_{[A}\delta^C_{B]}/rY. \tag{A6}$$

Using as our set U the union of the two spherical regions on each sheet given by $1 \leq r \leq r_0$ (see Fig. 4), we find that γ^C_{AB} and, therefore, C^C_{AB} are zero. The differential equation now reduces to

$$\nabla \cdot (e^A \gamma^C_{AB}) \epsilon_{BCM} = 0 \tag{A7}$$

or

$$\frac{\partial}{\partial x^a}\left\{\frac{Y^3}{r^3}\left[\frac{r}{Y}\delta^{aA}\right]\left(-\frac{1}{rY}\right)(x_A\delta^C_B - x_B\delta^C_A)\right\}\epsilon_{BCM} = 0. \tag{A8}$$

The boundary condition becomes

$$\frac{1}{V}\int_{\partial U}(x_A\delta^C_B - x_B\delta^C_A)\epsilon_{BCM}\delta\theta^M\delta^a_A g_{ab}\, dS^b = 0, \tag{A9}$$

where

$$g_{ab}\, dS^b = x_a(Y^3/r^2)\sin\theta\, d\theta\, d\phi. \tag{A10}$$

Both (A8) and (A9) are identically satisfied. Thus the space can be classified GBT I with the triad in (A5).

Concern about the continuity of the triad vectors at the surface where the two sheets have been identified may be removed by transforming the triad back to the Kantowski—Sachs coordinates and confirming that the triad is well-behaved at $\rho = 0$.

We may also calculate the value of I for this solution. Since

$$\Delta^C_{AB}\Delta^C_{AB}=\gamma^C_{AB}\gamma^C_{AB}=4/Y^2, \tag{A11}$$

we find that

$$I=2\lim_{r_0\to\infty}(1/V)\int(4/Y^2)\,dV$$
$$=8/Y^2. \tag{A12}$$

It seems likely that the symmetry group is responsible for Δ^2 being constant. Nevertheless, I is nonzero because the group is not simply transitive.

APPENDIX B: A COMPENDIUM OF THE GOWDY SOLUTION[24]

The metric is

$$ds^2=A^2(-dt^2+d\theta^2)+2t(D^2d\sigma^2+D^{-2}d\delta^2), \tag{B1}$$

where

$$A=(2t)^{-1/2}\exp(-3B_+/2),\quad D=\exp(\sqrt{3}B_-), \tag{B2}$$

and

$$0\leq\theta\leq 2\pi,\quad \oint\!\!\oint d\sigma\,d\delta=16\pi. \tag{B3}$$

The metric variables are of the form

$$B_-=\tfrac{1}{2}a_0\ln 2t+\sum_{n=1}^{\infty}c_nZ_0(nt)\cos[n(\theta-\theta_n)], \tag{B4}$$

$$B_+=-\tfrac{1}{2}a_0^2\ln 2t-\tfrac{1}{2}\sum_{n=1}^{\infty}c_n^2n^2\int^t dt'$$
$$\times 2t'[Z_0^2(nt')+Z_1^2(nt')]-2a_0\sum_{n=1}^{\infty}c_nZ_0(nt)\cos[n(\theta-\theta_n)]$$
$$+\sum_{n=1}^{\infty}\sum_{\substack{m=1\\ m\neq n}}^{\infty}2tnmc_nc_mZ_0(mt)Z_1(nt)$$
$$\times\left\{\frac{\cos[n(\theta-\theta_n)+m(\theta-\theta_m)]}{n+m}-\frac{\cos[n(\theta-\theta_n)-m(\theta-\theta_m)]}{n-m}\right\}$$
$$+\sum_{n=1}^{\infty}nc_n^2tZ_0(nt)Z_1(mt)\cos[2n(\theta-\theta_n)], \tag{B5}$$

where a_0, c_n, and θ_n are constants, and Z_l is any real solution to the lth order Bessel equation. As $t\to 0$,

$$B_-\sim\tfrac{1}{2}a\ln 2t,\quad B_+\sim-\tfrac{1}{2}a^2\ln 2t, \tag{B6}$$

where a is defined by

$$a(\theta)\equiv a_0+2\sum_{n=1}^{\infty}c_n\cos[n(\theta-\theta_n)], \tag{B7}$$

so the metric behaves as

$$A\sim(2t)^{(3a^2-1)/4},\quad D\sim(2t)^{\sqrt{3}a/4}. \tag{B8}$$

In the other limit, as $t\to\infty$, we have

$$A\sim(2t)^{-1/4}\exp(-3B_+/2),\quad D\sim(2t)^{\sqrt{3}a_0/2}, \tag{B9}$$

$$B_+\sim-\tfrac{1}{2}a_0^2\ln 2t-\frac{2t}{\pi}\sum_{n=1}^{\infty}n(c_n^2+d_n^2), \tag{B10}$$

$$B_-\sim\tfrac{1}{2}a_0\ln 2t+2(\pi 2t)^{-1/2}$$
$$\times\sum_{n=1}^{\infty}n^{-1/2}\cos[n(\theta-\theta_n)][c_n\cos(nt-\pi/4)$$
$$+d_n\sin(nt-\pi/4)], \tag{B11}$$

where $d_n\equiv c_n\cdot$(ratio of regular to irregular Bessel functions).

APPENDIX C: GEODESIC-NORMAL HYPERSURFACES IN THE GOWDY METRIC

We wish to find a timelike unit vector $\mathbf{e}_0'$, tangent to geodesic curves. This vector is related to $\mathbf{e}_0$ and $\mathbf{e}_1$ of Sec. IV by a Lorentz transformation:

$$\mathbf{e}_0'=\mathbf{e}_0\cosh\psi(t,\theta)+\mathbf{e}_1\sinh\psi(t,\theta). \tag{C1}$$

The requirement that $\mathbf{e}_0'$ be geodesic reduces to a differential equation for Ψ:

$$\frac{\partial}{\partial t}[A\sinh\psi]=-\frac{\partial}{\partial\theta}[A\cosh\psi], \tag{C2}$$

which cannot, in general, be integrated analytically.

Assume that the hypersurfaces S_G', to which $\mathbf{e}_0'$ is normal, are given by

$$t=f(\theta,J). \tag{C3}$$

The unit normals, $\mathbf{n}$, to S_G' are then of the form

$$\mathbf{n}=[\mathbf{e}_0+\mathbf{e}_1\,\partial f/\partial\theta]/A[1-(\partial f/\partial\theta)^2]^{1/2}, \tag{C4}$$

which, when identified with (C1), gives

$$\cosh\psi=[1-(\partial f/\partial\theta)^2]^{-1/2} \tag{C5}$$

or, after solving for $\partial f/\partial\theta$, and using (C3),

$$\partial f/\partial\theta=\tanh\psi(f(\theta,J),\theta). \tag{C6}$$

Once ψ is known, this equation determines f.

The metric in these hypersurfaces is

$$ds'^2=[A^2(-dt^2+d\theta^2)+2t(D^2d\sigma^2+D^{-2}d\delta^2)]\big|_{t=f}$$
$$=A^2(f,\theta)[1-(\partial f/\partial\theta)^2]d\theta^2+2f\,[D^2(f,\theta)\,d\sigma^2$$
$$+D^{-2}(f,\theta)\,d\delta^2]. \tag{C7}$$

Finally, the differential volume element is

$$dV'=A(f,\theta)2t[1-(\partial f/\partial\theta)^2]^{1/2}d\theta\,d\sigma\,d\delta. \tag{C8}$$

when ψ and its derivatives are small, (C2) may be linearized and integrated to yield

$$\psi(t,\theta)\approx\frac{A(t_0,\theta)}{A(t,\theta)}\psi(t_0,\theta)$$
$$-\frac{1}{A(t,\theta)}\int_{t_0}^{t}dt'\frac{\partial}{\partial\theta}A(t',\theta). \tag{C9}$$

In Sec. IV we consider metrics satisfying $1-3a^2>0$, so that S_G' may be chosen to coincide with S_G as $t\to 0$. This necessitates choosing $t_0=0$ and $\psi(t_0,\theta)=0$. Therefore, to first order in ψ, S_G' in Sec. IV is given by $t=f(\theta,J)$, where

$$\frac{\partial f}{\partial\theta}\cong\psi(t,\theta)$$
$$\cong-A^{-1}(t,\theta)\int_0^t dt'\,\frac{\partial}{\partial\theta}A(t',\theta). \tag{C10}$$

[1]A. Spero and R. Baierlein, J. Math. Phys. 18, 1330 (1977).
[2]M.A.H. MacCallum, "Cosmological Models from a Geometric Point of View," in *Cargese Lectures in Physics*, *Vol.* 6, edited by E. Schatzman (Gordon and Breach, New York, 1973).

[3]R.H. Gowdy, Ann. Phys. (N.Y.) **83**, 203 (1974).
[4]R. Kantowski, Ph.D. thesis (University of Texas, Austin, 1966); R. Kantowski and R.K. Sachs, J. Math. Phys. **7**, 443 (1966).
[5]Our conventions are these: Upper case Latin indices designate members of the orthonormal set of vectors; lower case Latin indices designate coordinate components; ∂_a represents partial differentiation and ∇_a covariant differentiation; $\pounds_{\mathbf{A}}\mathbf{B} \equiv (\mathbf{A}\cdot\boldsymbol{\partial})\mathbf{B} - (\mathbf{B}\cdot\boldsymbol{\partial})\mathbf{A}$ is the Lie derivative of B with respect to A; square brackets on indices denotes antisymmetrization. Due to the orthonormality of the $\{\mathbf{e}_A\}$, the vector indices are raised and lowered with the Kronecker delta:

$$\mathbf{e}^A = \Sigma\,\delta^{AB}\mathbf{e}_B.$$

Unless otherwise stated, we sum over any pair of repeated vector indices in a product. For example,

$$\Delta^C_{AB}\Delta^C_{AB} \equiv \sum_{A,B,C=1}^{N} \Delta^C_{AB}\Delta^C_{AB}.$$

The Einstein summation convention is used for coordinate indices.

[6]See Ref. 1 for details.
[7]Orthonormal vectors which generate simply-transitive Lie groups are denoted by $\mathbf{E}_A$, to distinguish them from the orthonormal vectors $\mathbf{e}_A$ that do not necessarily generate a group.
[8]L.P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1923).
[9]J.J. Stoker, *Differential Geometry* (Wiley-Interscience, New York, 1969); J.A. Wolf, *Spaces of Constant Curvature* (Publish or Perish Press, Boston, 1974).
[10]K. Yano and S. Bocher, *Curvature and Betti Numbers*, Annals of Mathematics Studies No. 32 (Princeton U.P., Princeton, N.J., 1953).
[11]The general problem of finding the possible topologies of all two-dimensional spaces of constant curvature is called the Clifford—Klein space-form problem, and has been solved for all but the negative curvature spaces. Of the compact topologies, only those with an Euler characteristic of zero can admit global, Killing vector fields and be homogeneous. The curvature—Integra theorem requires that these compact, homogeneous spaces have zero curvature. See Ref. 9.
[12]Aside from the Gowdy spacetime, we have analyzed one other known inhomogeneous solution to the Einstein equations: the ELS metrics [D. Eardley, E. Liang, and R.K. Sachs, J. Math. Phys. **13**, 99 (1972)]. They were found to have an approximate symmetry group of types I or V, depending upon the asymptotic behavior of the metric coefficients. This analysis has not been included here, for two reasons: (i) As an example, the ELS metrics illustrate nothing not shown in the Gowdy example; (ii) the metrics are not known globally, and thus inessential complications arise in the analysis. See A. Spero, Ph.D thesis (Wesleyan University, Middletown, CT, 1976).
[13]These spacetimes are actually Einstein—Rosen cylindrical-wave solutions that have been globally modified to make them compact. See R.H. Gowdy, Ann. Phys. (N.Y.) **83**, 203 (1974).
[14]B.K. Berger, Ann. Phys. (N.Y.) **83**, 458 (1974).
[15]A geodesic-normal family of hypersurfaces need not exist for every one of the Gowdy metrics.
[16]Triads other than (4.8) can be found which satisfy the differential equation and boundary condition. Consider $\{\mathbf{e}'_A\}$ such that $\mathbf{e}'_A$ commutes with the Killing vectors and $\mathbf{e}'_2$ and $\mathbf{e}'_3$ line in the group orbits. This gives a one-parameter family of triads, representing a "screw motion" in the θ direction, which includes (4.8) and solves the variational problem. For this family: $\eta^{23}, \eta^{22}, \eta^{33}$, and α^1 are nonzero; $\bar{\eta}^{AB}$ and $\bar{\alpha}^B$ are zero. The value of I for these solutions is demonstrably larger than I for (4.8); hence we shall not consider these solutions further. See A. Spero, Ph.D. thesis (Ref. 12).
[17]R.A. Issacson, Phys. Rev. **166**, 1263, 1272 (1968).
[18]In the intermediate case, where $1 - 3a^2 = 0$, $a(\theta)$ is a constant and the spacetime metric is exactly spatially-homogeneous.
[19]We can always express α as the divergence of a vector[1]; then Green's theorem implies that $\bar{\alpha}_A{}^A = 0$ for all compact, orientable spaces.
[20]Since I measures the amount of intrinsic inhomogeneity (in a normalized fashion), it is the appropriate quantity to use here.
[21]Although the normalized derivative tends to zero, both Θ_{AB} and $\pounds_{\mathbf{e}_L}\Theta_{AB}\,\delta_{L1}\delta_{AB}(2t)^{-(1+3a^2)/2}\ln(2t)\cdot(\text{function of }\Theta)$ go to infinity as $t \to 0$.
[22]The triad components of the Ricci tensor can be computed without knowing the metric explicitly because the geometry of the approximant is determined by the symmetry group, specifically, by the structure constants of an orthonormal triad.
[23]P. Szekeres, Comm. Math. Phys. **41**, 55 (1975).
[24]This material is adapted from Berger, Ref. 14.

Nonlinear realizations of the direct product of two Lorentz groups on a skew-symmetric tensor space

B. J. Dalton

Ames Laboratory-USDOE and Department of Physics, Iowa State University, Ames, Iowa 50011
(Received 14 November 1977)

In this paper, we present a study of exact, infinitesimal, nonlinear realizations of the direct product $(L_1\times L_2)$ of two Lorentz groups on an antisymmetric tensor space η. Included in this study is a detailed discussion of the second order $(L_1\times L_2)/L^+$ coset realizations where L^+ is the symmetric Lorentz subgroup. Invariant metric forms defined on η are exhibited and compared with invariant metric forms obtained in some previous studies of coset realizations.

I. INTRODUCTION

Direct product of homogeneous Lorentz groups (or their covering groups) have been studied previously in conjunction with analyticity problems of scattering amplitudes[1] and in composite particle models involving relativistic internal symmetry.[2,3]

In the present work we study the direct product $(L_1\times L_2)$ of two Lorentz groups realized as a transformation group on a space (η) of six antisymmetric tensor components. We also discuss the second order $(L_1\times L_2)/L^+$ coset realizations and show that these are contained in the above-mentioned realizations. In these realizations the transformations on η from the complement of L^+ in $L_1\times L_2$ are nonlinear.

The particular realizations of $L_1\times L_2$ studied here have possible uses within the context of Refs. 1, 2, and 3. In addition, there is the interesting possibility that physical fields might exist which transform nonlinearly under the Lorentz group.[4,5] In the particular realizations considered here, both Lorentz subgroups L_1 and L_2 act separately as nonlinear transformation groups on η. Lagrangians constructed from the $L_1\times L_2$ invariant metrics defined on η are invariant under the separate action of L_1 and L_2. The development in this paper is limited to discussion of the forms for these nonlinear realizations and to construction of the invariant metrics.

In Sec. II we shall establish the notation to be used and define the group generators. These are used in Sec. III where we give exact infinitesimal realizations of the group. In Sec. IV we compare these with coset realizations and, finally, in Sec. V we shall discuss several invariant metric forms.

II. NOTATION AND THE GROUP ALGEBRA

Here, the notation used in this paper is established, and a generator basis that will be convenient for the algebraic operations to follow is described. For further convenience in manipulation, the Pauli metric convention of letting the fourth space–time component be imaginary is used (that is, $x_4=ict$), together with the convention of summing from 1 to 4 (from 1 to 3) over repeated-Greek (Latin) indices.

The following generator basis is first introduced

$$t^{\pm}_{\mu\nu}=j_{\mu\nu}(1)\pm j_{\mu\nu}(2). \tag{2.1}$$

Here, the $j_{\mu\nu}(1)$ and $j_{\mu\nu}(2)$ are generators for the Lie algebras $\underset{\sim}{L}_1$ and $\underset{\sim}{L}_2$ of the respective Lorentz groups L_1 and L_2. Operators with index (1) commute with operators with index (2). The operators $t^{\pm}_{\mu\nu}$ satisfy the following commutation relations:

$$\begin{aligned}[t^-_{\mu\nu},t^-_{\rho\alpha}]&=[t^+_{\mu\nu},t^+_{\rho\alpha}]\\&=i(\delta_{\mu\rho}t^+_{\nu\alpha}-\delta_{\mu\alpha}t^+_{\nu\rho}+\delta_{\nu\alpha}t^+_{\mu\rho}-\delta_{\nu\rho}t^+_{\mu\alpha}),\\[t^-_{\mu\nu},t^+_{\rho\alpha}]&=i(\delta_{\mu\rho}t^-_{\nu\alpha}-\delta_{\mu\alpha}t^-_{\nu\rho}+\delta_{\nu\alpha}t^-_{\mu\rho}-\delta_{\nu\rho}t^-_{\mu\alpha}).\end{aligned} \tag{2.2}$$

From these relations, it is obvious that the set $k^+\equiv\{t^+_{\mu\nu}\}$ generates a Lorentz subalgebra. The set k^+, together with its complement part, $k^-\equiv\{t^-_{\mu\nu}\}$ comprise a generator basis for the Lie algebra of $L_1\times L_2$. It should be noted that k^- itself is not a subalgebra.

For the manipulation to follow, it is convenient to use a vector form for the operators. The vector operators corresponding to the above tensor operators are defined as follows:

$$J^{\pm}_i=\frac{i\epsilon_{ijk}t^{\pm}_{jk}}{2}\,,\quad K^{\pm}_i=t^{\pm}_{4i} \tag{2.3}$$

where ϵ_{ijk} is the usual totally antisymmetric tensor with normalization $\epsilon_{123}=+1$. These vector operators, as defined here for these finite nonunitary realizations, are real and satisfy the following commutation relations:

$$\begin{aligned}&[J^-_i,K^-_j]=[J^+_i,K^+_j]=-\epsilon_{ijk}K^+_k,\\&[K^-_i,K^-_j]=[K^+_i,K^+_j]=+\epsilon_{ijk}J^+_k,\\&[J^-_i,J^-_j]=[J^+_i,J^+_j]=-\epsilon_{ijk}J^+_k,\\&[J^-_i,K^+_j]=-\epsilon_{ijk}K^-_k,\quad[K^-_i,K^+_j]=+\epsilon_{ijk}J^-_k,\\&[J^-_i,J^+_j]=-\epsilon_{ijk}J^-_k.\end{aligned} \tag{2.4}$$

In terms of these operators, a second set of basis operators is defined as follows:

$$T^{\pm}_k=\tfrac{1}{2}(J^{\pm}_k+iK^{\pm}_k),\quad(T^{\pm}_k)^*=\tfrac{1}{2}(J^{\pm}_k-iK^{\pm}_k). \tag{2.5}$$

The commutation relations for these operators are given by

$$\begin{aligned}&[T^+_i,T^+_j]=[T^-_i,T^-_j]=-\epsilon_{ijk}T^+_k,\\&[T^-_i,T^+_j]=-\epsilon_{ijk}T^-_k,\quad[T^{\pm}_i,(T^{\pm}_j)^*]=0\end{aligned} \tag{2.6}$$

with similar relations holding for the complex conjugate components. The sets $T=\{T^+_i,T^-_i\}$ and $T^*=\{(T^+_i)^*,(T^-_i)^*\}$ together form a generator basis for the algebra

 0022-2488/78/1906-1335$1.00

of $L_1\times L_2$. Furthermore, since the operators in T commute with those in T^*, the above basis corresponds to an alternate direct product decomposition (indicated by $T_1\times T_2$) of the general group $L_1\times L_2$. One convenient feature of the latter basis is that the generators $\{T^+, (T^+)^*\}$ generate the subgroup L^+.

The connections between the infinitesimal group parameters corresponding to the above operator sets are established by the following relations:

$$\begin{aligned}\tfrac{1}{2}i[\omega_{\mu\nu}(1)\,j_{\mu\nu}(1)+\omega_{\mu\nu}(2)\,j_{\mu\nu}(2)] &= \tfrac{1}{2}i(\omega^+_{\mu\nu}t^+_{\mu\nu}+\omega^-_{\mu\nu}t^-_{\mu\nu})\\ &= \boldsymbol{\alpha}^+\cdot\mathbf{J}^+ - \boldsymbol{\nu}^+\cdot\mathbf{K}^+ + \boldsymbol{\alpha}^-\cdot\mathbf{J}^- - \boldsymbol{\nu}^-\cdot\mathbf{K}^+\\ &= \boldsymbol{\beta}^+\cdot\mathbf{T}^+ + (\boldsymbol{\beta}^+)^*\cdot(\mathbf{T}^+)^* + \boldsymbol{\beta}^-\cdot\mathbf{T}^- + (\boldsymbol{\beta}^-)^*\cdot(\mathbf{T}^-)^*. \end{aligned}\tag{2.7}$$

From these relations and the previous relations between the different sets of generators, the explicit connections between infinitesimal parameters are determined as follows:

$$\left.\begin{aligned}\omega_{\mu\nu}(1) &= \omega^+_{\mu\nu}+\omega^-_{\mu\nu}\\ \omega_{\mu\nu}(2) &= \omega^+_{\mu\nu}-\omega^-_{\mu\nu}\end{aligned}\right\},\quad \omega^\pm_{\mu\nu}=\frac{\omega_{\mu\nu}(1)\pm\omega_{\mu\nu}(2)}{2},\tag{2.8}$$

$$\begin{aligned}&\alpha^\pm_i=\tfrac{1}{2}\epsilon_{ijk}\omega^\pm_{jk},\quad \omega^\pm_{ij}=\epsilon_{ijk}\alpha^\pm_k,\\ &\nu^\pm_i=-i\omega^\pm_{4i},\quad \omega^\pm_{4i}=+i\nu^\pm_i,\end{aligned}\tag{2.9}$$

$$\begin{aligned}&\beta^\pm_i=\alpha^\pm_i+i\nu^\pm_i,\quad \alpha^\pm_i=[\beta^\pm_i+(\beta^\pm_i)^*]/2,\\ &(\beta^\pm_i)^*=\alpha^\pm_i-i\nu^\pm_i,\quad \nu^\pm_i=[\beta^\pm_i-(\beta^\pm_i)^*]/2.\end{aligned}\tag{2.10}$$

III. EXACT INFINITESIMAL REALIZATIONS

Here, a general infinitesimal $L_1\times L_2$ group action on the six components $(\eta_{\mu\nu})$ of an antisymmetric tensor is considered. In this group action, the subgroup (L^+) is assumed to act on the space η in a linear way. To obtain the infinitesimal realizations of the L^- part of the group action on this space, a procedure used by Weinberg[6] in his analysis of the nonlinear realizations of the chiral group SU(2)×SU(2) is followed. This procedure involves the assumption of a general form for the L^- action, and then the restriction of this form by imposition of the commutation relations of the Lie algebra.

The linear action of the subgroup L^+ on the field tensors $\eta^{\mu\nu}$ is given by the following expression:

$$\eta^{\mu\nu}=(\Lambda^\mu{}_\rho)^+(\Lambda^\nu{}_\beta)^+\eta^{\rho\beta}\tag{3.1}$$

with $\Lambda^\mu{}_\nu=\delta^\mu{}_\nu+\omega^\mu{}_\nu$; the infinitesimal realizations of L^+ are given by

$$\delta\eta^{\mu\nu}=(\omega^\nu{}_\beta)^+\eta^{\mu\beta}-(\omega^\mu{}_\beta)^+\eta^{\nu\beta}.\tag{3.2}$$

With the $T\times T^*$ decomposition it is more convenient to work with two complex field vectors defined as follows:

$$\begin{aligned}&\mathbf{S}=\mathbf{M}+i\mathbf{V},\quad \mathbf{M}=(\mathbf{S}+\mathbf{S}^*)/2,\\ &\mathbf{S}^*=\mathbf{M}+i\mathbf{V},\quad \mathbf{V}=(\mathbf{S}-\mathbf{S}^*)/2i,\end{aligned}\tag{3.3}$$

where

$$M^i=\epsilon^{ijk}\eta^{jk}/2,\quad V^i=-i\eta^{4i}.\tag{3.4}$$

The group realizations on $\mathbf{M}$ and $\mathbf{V}$ can be found from the group action on $\mathbf{S}$ and $\mathbf{S}^*$ by using (3.3). From (3.2), the L^+ group action on $\mathbf{S}^*$ and $\mathbf{S}$ is given by the following equations:

$$\delta S_i=-\epsilon_{ijk}\beta^+_j S_k,\quad \delta S^*_i=-\epsilon_{ijk}(\beta^+_j)^*S^*_k.\tag{3.5}$$

From this point on, it is necessary to write only the group action on the complex vector $\mathbf{S}$. The group action on $\mathbf{S}^*$ is obtained simply by complex conjugation of the relations (including the group parameters) for the group action on $\mathbf{S}$.

For the parameter β^+_j, the action of the group algebra is given by the relation

$$[T^+_j,S_i]=-\frac{\partial S'_i}{\partial\beta^+_j}=+\epsilon_{ijk}S_k=-\epsilon_{jik}S_k.\tag{3.6}$$

Here $S'_i\equiv S_i+\delta S_i$. For the action of T^- we assume the following form involving two undetermined functions G and F.

$$[T^-_j,S_i]=-\frac{\partial S'_i}{\partial\beta^-_j}=G\delta_{ij}+FS_iS_j\equiv A_{ij}.\tag{3.7}$$

In the operations to follow, it is assumed that the noninvariant dependence of the functions G and F upon the field arises only through the function $S^2=S_iS^i$ which is invariant under the subgroup L^+ (notice that $S^2=\mathbf{M}^2-\mathbf{V}^2+2i\mathbf{V}\cdot\mathbf{M}$). These two functions ($G$ and F) may also depend upon any other $L_1\times L_2$ invariant function, say D, where $[T^\pm_i,D]=0$. This latter dependence is discussed in Sec. V in terms of invariant forms, but because $[T^\pm,D]=0$, it will not be involved in the commutation condition considered below.

The above infinitesimal group action on the complex 3-vector $\mathbf{S}$ is next required to satisfy the following commutation relation:

$$[T^-_i,[T^-_i,S_j]]-[T^-_i,[T^-_i,S_j]]=[[T^-_i,T^-_i],S_j].\tag{3.8}$$

Using the explicit commutation relations (2.6), this requirement, together with the following formula,

$$[T^-_i,G]=2G'[G+FS^2]S_i,\quad G'=\frac{dG}{d(S^2)},\tag{3.9}$$

which follows from (3.7), leads directly to the following first-order differential equation

$$GF-2G'(G+FS^2)=1.\tag{3.10}$$

The functions F and G are arbitrary subject to this equation. Under a redefinition of the space of the form

$$S_i\to\bar{S}_i=\phi S_i\tag{3.11}$$

one realization may be transformed into another if the function ϕ, $\bar{F}$ and $\bar{G}$ satisfy the following conditions

$$\bar{G}=\phi G,\tag{3.12}$$

$$\bar{F}=\phi^{-2}[\phi F+2\phi'(G+FS^2)].\tag{3.13}$$

For use in Sec. V, the following realization form is discussed here:

$$G=(D^2-S^2)^{1/2}.\tag{3.14}$$

Here, D is any $L_1\times L_2$ invariant. With (3.14), Eq. (3.10) reduces to

$$(G+S^2/G)F=0.\tag{3.15}$$

One has two possible cases which satisfy the zero product.

$$D\neq 0,\quad G=(D^2-S^2)^{1/2},\quad F=0,\tag{3.16}$$

$$D=0, \quad G=(-S^2)^{1/2}, \quad F=\text{arbitrary function.} \tag{3.17}$$

The $L_1\times L_2$ invariant forms for these two realizations are discussed in Sec. V with a particular choice for F in the second case.

Using the relations between the group parameters, the nonlinear group action on the M_i and V_i vector components may be expressed in the following form:

$$[J_k^-, M_l]=-\frac{\partial M_l'}{\partial \alpha_k}=\tfrac{1}{2}(A_{kl}+A_{kl}^*), \tag{3.18}$$

$$[J_k^-, V_l]=-\frac{\partial V_l'}{\partial \alpha_k}=-\frac{i}{2}(A_{kl}-A_{kl}^*), \tag{3.19}$$

$$[K_k^-, M_l]=+\frac{\partial M_l'}{\partial \nu_k}=-\frac{i}{2}(A_{kl}-A_{kl}^*), \tag{3.20}$$

$$[K_k^-, V_l]=+\frac{\partial V_l'}{\partial \nu_k}=-\tfrac{1}{2}(A_{kl}+A_{kl}^*). \tag{3.21}$$

In the above realizations of $L_1\times L_2$, both L_1 and L_2 act as nonlinear transformation groups on η. Consider the following basis for L_1 and L_2:

$$T_k(i)=\tfrac{1}{2}[J_k(i)+iK_k(i)], \quad T_k^*(i)=\tfrac{1}{2}[J_k(i)-iK_k(i)], \quad i=1,2, \tag{3.22}$$

where $J_k(i)$ and $K_k(i)$ are defined as in (2.3). From (2.1), (2.3), and (2.5) one has the following relations:

$$T_k(i)=\frac{T_k^+ + A_i T_k^-}{2}, \quad T_k^*(i)=\frac{(T_k^+)^* + A_i (T_k^-)^*}{2}, \tag{3.23}$$

where $A_i=-(-1)^i$, $i=1,2,$. Equations (3.6) and (3.7) then give the following expressions for $T_j(1)$ of L_1:

$$[T_j(1), S_i]=-\tfrac{1}{2}\epsilon_{jik}S_k+\tfrac{1}{2}G\delta_{ij}+\tfrac{1}{2}FS_iS_j. \tag{3.24}$$

With (3.24) and its complex conjugation one has a well-defined nonlinear infinitisimal realization of L_1 on the space η. For L_2 one has the following relation and its complex conjugate.

$$[T_j(2), S_i]=-\tfrac{1}{2}\epsilon_{jik}S_k-\tfrac{1}{2}G\delta_{ij}-\tfrac{1}{2}FS_iS_j. \tag{3.25}$$

It is clear from (3.24) and (3.25) that the realizations of L_1 and L_2 on η can be obtained one from the other by the inversion $G\to -G$ and $F\to -F$.

IV. COMPARISON WITH COSET REALIZATIONS

In this section, the infinitesimal group realizations on the coset space of $(T\times T^*)/T^*$ are considered, where $T\times T^*$ is the alternate direct product decomposition of $L_1\times L_2$ mentioned in Sec. II. These coset realizations through second order are compared with corresponding second order expansions of the above exact realizations for a particular choice of the classifying function G.

Following the work of earlier authors,[4,5,7-9] the exponential map is used to express a general group element of $T\times T^*$, that is,

$$g=\exp[\mathbf{S}\cdot\mathbf{T}^- + \mathbf{S}^*\cdot(\mathbf{T}^-)^*]\exp[\mathbf{h}\cdot\mathbf{T}^+ + \mathbf{h}^*\cdot(\mathbf{T}^+)^*]. \tag{4.1}$$

Since the generators $\{T_i^+, (T_j^+)^*\}$ generate the subgroup L^+, the particular factoring in (4.1) corresponds to the coset space $(L_1\times L_2)/L^+$ which can be parameterized by six antisymmetric tensors. Since the T and T^* parts commute with each other, Eq. (4.1) may be rewritten in the following form:

$$g=\exp[\mathbf{S}\cdot\mathbf{T}^-]\exp[\mathbf{h}\cdot\mathbf{T}^+]\exp[\mathbf{S}^*\cdot(\mathbf{T}^-)^*]\exp[\mathbf{h}^*\cdot(\mathbf{T}^+)^*]. \tag{4.2}$$

From this form it can be seen that it is necessary to consider only the T group action since the T^* group action follows as in Sec. III by complex conjugation. The action of an arbitrary group element $g_0=\exp[\boldsymbol{\beta}^-\cdot\mathbf{T}^-]\times\exp[\boldsymbol{\beta}^+\cdot\mathbf{T}^+]$ on this coset space is given as follows:

$$g_0\exp[\mathbf{S}\circ\mathbf{T}^-]=\exp(\mathbf{S}'\circ\mathbf{T}^-)\exp(\mathbf{u}\circ\mathbf{T}^+). \tag{4.3}$$

The following expansions are for terms in $\mathbf{u}$ and $\delta\mathbf{S}=\mathbf{S}'-\mathbf{S}$ through second order in S and through first order in the parameters of g_0. The explicit second-order realizations are found by using the following Baker-Campbell-Hausdorff (BCH)[10] expansion formula:

$$\exp(a)\exp(b)=\exp(a+b+\tfrac{1}{2}[a,b]+\tfrac{1}{12}[[a,b],b] - \tfrac{1}{12}[[a,b],a]+\cdots). \tag{4.4}$$

Setting $g_0=\exp(\boldsymbol{\beta}^-\cdot\mathbf{T}^-)$, both sides of (4.3) are expanded by (4.4) giving the following relation:

$$\boldsymbol{\beta}^-\cdot\mathbf{T}^- + \mathbf{S}\cdot\mathbf{T}^- + \tfrac{1}{2}[\boldsymbol{\beta}^-\circ\mathbf{T}^-, \mathbf{S}\circ\mathbf{T}^-] + \tfrac{1}{12}[[\boldsymbol{\beta}^-\circ\mathbf{T}^-, \mathbf{S}\circ\mathbf{T}^-], \mathbf{S}\cdot\mathbf{T}^-]$$
$$= \mathbf{S}'\cdot\mathbf{T}^- + \mathbf{u}\circ\mathbf{T}^+ + \tfrac{1}{2}[\mathbf{S}'\cdot\mathbf{T}^-, \mathbf{u}\cdot\mathbf{T}^+] - \tfrac{1}{12}[[\mathbf{S}'\cdot\mathbf{T}^-, \mathbf{u}\cdot\mathbf{T}^+], \mathbf{S}'\circ\mathbf{T}^-]. \tag{4.5}$$

Since the generators of T^+ and T^- are linearly independent, this relation reduces to the following two equations:

$$\mathbf{u}\cdot\mathbf{T}^+ = \tfrac{1}{2}[\boldsymbol{\beta}^-\cdot\mathbf{T}^-, \mathbf{S}\cdot\mathbf{T}^-] + \tfrac{1}{12}[[\mathbf{S}\circ\mathbf{T}^-, \mathbf{u}\cdot\mathbf{T}^+], \mathbf{S}\circ\mathbf{T}^-], \tag{4.6}$$

$$\delta\mathbf{S}\cdot\mathbf{T} = \boldsymbol{\beta}^-\cdot\mathbf{T}^- - \tfrac{1}{2}[\mathbf{S}\circ\mathbf{T}^-, \mathbf{u}\cdot\mathbf{T}^+] + \tfrac{1}{12}[[\boldsymbol{\beta}^-\circ\mathbf{T}^-, \mathbf{S}\circ\mathbf{T}^-], \mathbf{S}\circ\mathbf{T}^-]. \tag{4.7}$$

Iterating (4.6) and using the result in (4.7), the second-order solutions become

$$\mathbf{u}\cdot\mathbf{T}^+ = \tfrac{1}{2}[\boldsymbol{\beta}^-\circ\mathbf{T}^-, \mathbf{S}\cdot\mathbf{T}^-] + O(S^3), \tag{4.8}$$

$$\delta\mathbf{S}\cdot\mathbf{T}^- = \boldsymbol{\beta}^-\cdot\mathbf{T}^- + \tfrac{1}{3}[[\boldsymbol{\beta}^-\cdot\mathbf{T}^-, \mathbf{S}\cdot\mathbf{T}^-], \mathbf{S}\circ\mathbf{T}^-] + O(S^3). \tag{4.9}$$

The above method for obtaining these expansion formulas was outlined in the recent work of Machacek and McCliment.[11] The form of formula (4.9) was also obtained by Hopkinson and Reya[5] for the $O(3,1)$ realizations on the coset space of $O(3,1)/O(3)$ via an alternate technique.

For $g_0=\exp(\boldsymbol{\beta}^+\cdot\mathbf{T}^+)$, a repeat of the steps in the above procedure gives the following expressions:

$$\mathbf{u}\cdot\mathbf{T}^+ = \boldsymbol{\beta}^+\circ\mathbf{T}^+, \tag{4.10}$$

$$\delta\mathbf{S}\circ\mathbf{T}^- = [\boldsymbol{\beta}^+\cdot\mathbf{T}^+, \mathbf{S}\cdot\mathbf{T}^-]. \tag{4.11}$$

As can be seen the subgroup T^+ acts linearly on the coset coordinates with $\mathbf{u}=\boldsymbol{\beta}^+$.

Using the explicit commutation relations described in Sec. II, the expressions in (4.9) and (4.11) take on the following form:

$$\frac{\partial S_i'}{\partial \beta_j^-} = (1 - \tfrac{1}{3}S^2)\,\delta_{ij} + \tfrac{1}{3}S_iS_j + O(S^3), \tag{4.12}$$

$$\frac{\partial S_i'}{\partial \beta_j^+} = -\epsilon_{ijk}S_k. \tag{4.13}$$

Equation (4.13) is exact and corresponds to the linear action described by Eq. (3.6). To compare the second-order coset realizations in (4.12) with the exact realizations discussed in Sec. III, it is necessary to find an appropriate form for the classifying function G and then make second-order expansions of Eq. (3.7).

To describe (4.12) by (3.7), G must reduce to a second order polynomial and F to a constant in the second order expansion. It is possible to satisfy (3.10) with $G = aS^2 + b$ and $F = K = \text{const}$ if $a = -K/2$ and $B = 1/2K$. This realization, however, will not satisfy the coset realizations given in (4.12) since $K = 1/3$ will give coefficients of δ_{ij} which do not agree with the corresponding turn in (4.12). However, it is possible to transform, as in (3.11), to a new set of variables for which these L^- realizations reduce, in second order expansions, to the coset realizations. With the above mentioned choice of G and F consider the variable transformation

$$S_i = h(\rho)\rho_i, \quad h(\rho^2) = 1 + a\rho^2. \tag{4.14}$$

The realization of the ρ space has the form

$$-\frac{\partial \rho_i'}{\partial \beta_j^-} = \frac{1}{h}\left(\frac{1}{2K} - \frac{Kh^2\rho^2}{2}\right)\delta_{ij} + \frac{1}{K}\left(\frac{K^2h^3 - a + aK^2h^2\rho^2}{h^2 + 2a\rho^2 h}\right)\rho_i\rho_j. \tag{4.15}$$

This realization reduces in second order expansions to the coset realizations in (4.12) if a and K satisfy the conditions

$$1 = -2K, \quad \text{and} \quad 3K^2 + K - 3a = 0. \tag{4.16}$$

These equations are satisfied for $K = -1/2$ and $a = 1/12$. The above realizations on the $\mathbf{S}$ space with $K = -1/2$ have the form

$$-\frac{\partial S_i'}{\partial \beta_j^-} = \left(\frac{S^2}{4} - 1\right)\delta_{ij} - \tfrac{1}{2}S_iS_j. \tag{4.17}$$

This particular realization is interesting because, as shown in the following section, it leaves invariant the diagonal conformally flat de Sitter metric defined on the $\mathbf{S}$ space. The transformation $S_i = (1 + \rho^2/12)\,\rho_i$ is then just the change of variables needed to diagonalize the invariant metric on the coset space.

With the above comparison, it is clear that for at least one choice of the function G the exact realizations of Sec. III have the same form as the above coset realizations, at least through second order. It should be clear that realizations with choices for G and F other than in (4.15) may exist which reduce upon second order expansions to the second order coset realizations.

V. INVARIANT FORMS AND LAGRANGIANS

In this section, several invariant forms defined on the vector components which span the tensor space are discussed. The invariance under $L_1 \times L_2$ of the L^+ invariant forms $\mathbf{M}^2 - \mathbf{V}^2$ and $\mathbf{M}\cdot\mathbf{V}$ is first discussed. It is shown that the invariance of these expressions requires a particular realization, that is, a particular choice for the function G. Invariant Riemann metrics on the tensor space are constructed and compared with invariant forms constructed for coset spaces in previous studies.[5,12]

Using the equation $\mathbf{S} = \mathbf{M} + i\mathbf{V}$ the usual L^+-invariant forms $\mathbf{M}^2 - \mathbf{V}^2$ and $\mathbf{M}\cdot\mathbf{V}$ can be rewritten as follows:

$$\mathbf{M}^2 - \mathbf{V}^2 = \frac{S^2 + (S^*)^2}{2}, \quad \mathbf{M}\cdot\mathbf{V} = \frac{S^2 - (S^*)^2}{4i}. \tag{5.1}$$

From the general L^- action defined in (3.7), one can obtain the following relations:

$$-\delta(\mathbf{M}^2 - \mathbf{V}^2) = [(G + FS^2)\,\mathbf{S}\cdot\boldsymbol{\beta}^-] + [(G + FS^2)\,\mathbf{S}\cdot\boldsymbol{\beta}^-]^*, \tag{5.2}$$

$$-\delta(\mathbf{M}\cdot\mathbf{V}) = (1/2i)\{[(G + FS^2)\,\mathbf{S}\cdot\boldsymbol{\beta}^-] - [(G + FS^2)\,\mathbf{S}\cdot\boldsymbol{\beta}^-]^*\}. \tag{5.3}$$

Both of these forms are obviously invariant under the L^- action if $\mathbf{S}\cdot\boldsymbol{\beta}^- = 0$. However, this condition does not hold for an arbitrary element of L^-, given an arbitrary point in the η space. Both quantities are invariant under L^- (and thus under $L_1 \times L_2$) if the following condition holds true:

$$G + FS^2 = 0. \tag{5.4}$$

This equation must be considered simultaneously with Eq. (3.10) which together require the following forms for G and F:

$$G = \mp iS, \quad F = \pm i/S. \tag{5.5}$$

This realization corresponds to the second realization [see Eq. (3.17)] mentioned in Sec. III with a specific choice for the function G. With these relations, the group action (3.7) takes on the form

$$[T_j^-, S_i] = -\frac{\partial S_i'}{\partial \beta_j^-} = \mp i\left(\delta_{ij}S - \frac{S_iS_j}{S}\right). \tag{5.6}$$

The above realization is obviously not valid for $S = 0$. This means that if $\mathbf{M}^2 - \mathbf{V}^2$ and $\mathbf{M}\cdot\mathbf{V}$ are to be $L_1 \times L_2$ invariants, then at least one of these expressions must differ from zero since $S^2 = \mathbf{M}^2 - \mathbf{V}^2 + 2i\mathbf{M}\cdot\mathbf{V}$.

Next, consider the realization given by (3.16) where $F = 0$ and $G = (D^2 - S^2)^{1/2}$, where D is some invariant function under $L_1 \times L_2$. The realizations (3.7) take on the form

$$[T_j^-, S_i] = -\frac{\partial S_i'}{\partial \beta_j^-} = (D^2 - S^2)^{1/2}\,\delta_{ij}. \tag{5.7}$$

Setting $S_4 = G$, one obtains the following relations:

$$-\frac{\partial S_i'}{\partial \beta_j^-} = +S_4\delta_{ij}, \quad -\frac{\partial S_4'}{\partial \beta_j^-} = -S_i\delta_{ij}. \tag{5.8}$$

These relations correspond to an infinitesimal rotation in planes which leaves the quantity D^2 invariant where D^2 satisfies

$$D^2 = S^2 + S_4^2. \tag{5.9}$$

It should be clear here that S_4 is a scalar under L^+ and

should not be confused with the fourth component of an L^+ 4-vector.

The complex conjugate of D is also invariant under $L_1 \times L_2$. With this, one can set $S_4 = \phi + i\psi$ and obtain the following two finite invariant forms corresponding to the real and imaginary parts of D:

$$D_1 = \mathbf{M}^2 - \mathbf{V}^2 + \phi^2 - \psi^2, \tag{5.10}$$

$$D_2 = 2\mathbf{M} \cdot \mathbf{V} + 2\phi\psi. \tag{5.11}$$

The functions ϕ and ψ are L^+ scalars. The invariance of the forms (5.10) and (5.11) depends upon the use of the realizations as given in (5.7).

In addition to the above forms, it is possible to construct an infinitesimal invariant form on the field. For the realization given in (5.7), the following Riemann metric form is invariant under $L_1 \times L_2$:

$$d\eta^2 = d\mathbf{S} \cdot d\mathbf{S} + (dS_4)^2. \tag{5.12}$$

For the particular realization where the invariant function D is constant, it is possible to construct an infinitesimal metric form that does not involve the scalar field components. This case corresponds to the realization for which $G = (1 - S^2)^{1/2}$, where for convenience D has been normalized to one. From the geometric point of view, the constant D is just the radius of curvature. With D constant ($D = 1$), the differential of both sides of Eq. (5.9) gives $\mathbf{S} \cdot d\mathbf{S} + S_4 dS_4 = 0$ from which one can obtain the relation

$$dS_4 = (\mathbf{S} \cdot d\mathbf{S})/(1 - S^2)^{1/2} \tag{5.13}$$

The use of (5.13) in (5.12) gives the metric form

$$d\eta^2 = d\mathbf{S} \cdot d\mathbf{S} + (\mathbf{S} \cdot d\mathbf{S})^2/(1 - S^2) \tag{5.14}$$

This particular metric form does not involve the scalar field components ϕ and ψ. It is by now a standard form appearing in several publications on coset realizations (see for instance, the work by Meetz[12] and references quoted therein) and in textbooks on general relativity (see for instance, Alder, Bazin, and Schiffer,[13] p. 348, from which the above procedure for deriving this metric form was obtained).

A worthwhile comparison can be made between the form in (5.14) with a particular metric form obtained by Hopkinson and Reya[5] in their study of the coset realizations of the Poincaré group. To make this comparison, first consider the following change of variables:

$$q_i = 2S_i/(1 + S_4), \quad S_4 \equiv (1 - S^2)^{1/2}. \tag{5.15}$$

In terms of these new variables, the metric in (5.14) takes on the following diagonal form:

$$d\eta^2 = Q^2 d\mathbf{q} \cdot d\mathbf{q}, \quad Q^{-1} = (1 + q^2/4). \tag{5.16}$$

This diagonal form is just the form of the well-known conformally flat de Sitter metric[14] defined here on a space of three complex variables. With the above variable relations, the L^- action in (5.7), expressed on the q space, is given by the following equation:

$$-\frac{\partial q_i'}{\partial \beta_j} = \left(\frac{q^2}{4} - 1\right)\delta_{ij} - \tfrac{1}{2} q_i q_j. \tag{5.17}$$

This equation is just the realization given in (4.17) with the substitution $S_i \to q_i$. Making the variable change $q_i = [1 + S^2/12]\rho_i$ discussed in Sec. III and then carrying out the second order expansion, the metric in (5.16) reduces to the form

$$dz^2 = (1 - \tfrac{1}{3}\rho^2)\, d\boldsymbol{\rho} \cdot d\boldsymbol{\rho} + \tfrac{1}{3}(\boldsymbol{\rho} \cdot d\boldsymbol{\rho})^2 + O(\rho^3). \tag{5.18}$$

This latter form is identical to that obtained by Hopkinson and Reya[5] [see Eq. (20) of this reference] in their study of the Poincaré group realizations on the coset space $O(3,1)/O(3)$.

In the above study we have discussed in some detail the mathematics of the $L_1 \times L_2$ group realizations on an antisymmetric tensor space. Several invariant metric forms and Lagrangians on this space were also discussed. The group $L_1 \times L_2$ is not a physical invariance group for a system of two interacting particles. However, by studying these coset type realizations we have obtained, although indirectly, nonlinear realizations of the Lorentz group (considered as L_1 or L_2) on a skew-symmetric tensor space η. The demonstrated fact that nonlinear realizations of the Lorentz group [or its SL(2, C) covering group] exist strongly suggests that such realizations might be constructed by a more direct method. It remains yet to be seen whether nonlinear realizations obtained by a more direct method will be limited to, differ from, or be more general than the ones obtained here. In any case the present work will be a worthwhile guide for future studies in this area.

ACKNOWLEDGMENTS

The author is grateful to Dr. E. Reya for checking a number of the mathematical results of this paper and for several helpful discussions on non-linear realizations. Dr. S. Williams and Dr. D. Pursey are acknowledged for making a critical reading of this manuscript.

[1] This problem is discussed by M. Toller and G. Domokos, in *Proceedings of the Eighth Nobel Symposium*, edited by N. Savartholm (Wiley, New York, 1968), pp. 15 and 33, respectively.
[2] T. Takabayasi, in *Proceedings of the* 1967 *International Conference on Particles and Fields*, edited by C.R. Hagen, G. Guralnik, and V.S. Mathur (Wiley, New York, 1967), p. 413.
[3] T. Takabayasi, in Ref. 1, p. 157, and references quoted therein.
[4] J.D. Hind, Nuovo Cimento A **4**, 71 (1971).
[5] J.F.L. Hopkinson and E. Reya, Phys. Rev. D **10**, 342 (1972).
[6] S. Weinberg, Phys. Rev. **166**, 1568 (1968).
[7] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 229 (1969).
[8] C.G. Collan, Jr. and S. Coleman, Phys. Rev. **177**, 2247 (1969).
[9] C.J. Isham, Nuovo Cimento A **59**, 356 (1969).
[10] The Baker–Campbell–Hausdorff (BCH) expansion formula is discussed in several textbooks on Lie algebras. See, for instance, N. Jacobson, *Lie Algebras* (Interscience, New York, 1966).
[11] M.E. Machacek and E.R. McCliment, Phys. Rev. D **10**, 1962 (1974).
[12] K. Meetz, J. Math. Phys. **10**, 589 (1969).
[13] K. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), p. 348.
[14] F. Gürsey, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1962), p. 365.

The distribution function for stretched dipoles in an applied electric field

D. P. Mason and D. K. McIlroy

Department of Applied Mathematics, University of the Witwatersand, Jan Smuts Avenue, Johannesburg, South Africa
(Received 9 November 1977)

An expression is derived for the distribution function of an assemblage of ion pairs in the presence of a uniform applied electric field, for the most interesting and important case, namely when the interaction force between the components of such stretched dipoles is unshielded Coulombic. Our derivation makes use of results obtained previously when determining, by means of a perturbation technique, the relative increase in the dissociation constant of a weak electrolyte due to an applied electric field. The expression we obtain for the ion-pair distribution function agrees with that given previously by Onsager [J. Chem. Phys. 2, 599 (1934)], whose derivation has never been published in full.

1. INTRODUCTION

The distribution function for an assemblage of rigid dipoles in a uniform applied electric field **X** is well known,[1] being simply the Maxwell–Boltzmann distribution. The determination of the distribution function $f(r, \theta)$ for an assemblage of associated ion pairs under Coulombic interaction in a uniform applied electric field where, of course, the distance r separating the ions is not constant (i.e., the problem of the stretched dipole) is a much more difficult problem, because, unlike the rigid dipole case, it involves nonequilibrium physics.

Onsager[2] has actually given a solution to this problem where it appears in consideration of the theory of Wien dissociation of weak[3] electrolytes. He found that

$$f(r, \theta) = (1/r)\exp[1/r + \epsilon r(\cos\theta - 1)] \times \int_0^1 J_0[(-8\epsilon s)^{1/2}\cos(\theta/2)]e^{-s/r}ds\,, \qquad (1.1)$$

where J_0 denotes the ordinary Bessel function of order zero and θ is the angle[4] an ion pair makes with the electric field **X**. The variable r has been made dimensionless on division by the characteristic length $2q$, where

$$q = -e_i e_j/2DkT > 0 \qquad (1.2)$$

is the Bjerrum association distance[5]; D is the dielectric constant of the solution, T its absolute temperature, k is Boltzmann's constant, and e_i and e_j are the charges of the ions. The distribution function has been renormalized by dividing by $n_i n_j$, where the constants n_i and n_j denote the probability densities of the "free" ions. Also $\epsilon = 2\beta q$, where

$$\beta = |X(e_i\omega_i - e_j\omega_j)|/2kT(\omega_i + \omega_j)$$

and ω_i and ω_j are the mobility coefficients of the ions. Onsager stated that the derivation of this important result "involves elaborate analysis," but did not publish his derivation in full.

Recently,[6] we gave a perturbation solution to the related problem of Wien dissociation in a weak electrolyte, using as perturbation parameter the dimensionless quantity $\epsilon = 2\beta q$. To calculate the relative increase in the dissociation constant $K(X)/K(0)$ due to the applied electric field **X**, we found that only the Legendre transform of the ion-pair distribution function, and not the distribution function itself, was required. In this paper we will show how our analysis can be extended to actually obtain an expression for the ion-pair distribution function. The result we obtain agrees with Eq. (1.1) given by Onsager.

It is possible that modifications of our perturbation technique may be used to determine the ion-pair distribution function in other problems in this and related fields. One generalization of interest would be to nonisotropic media which Onsager[7] states would be desirable but would lead to a more formidable partial differential equation for f. The experience gained here should help in the solution of these more general problems.

2. PARTIAL DIFFERENTIAL EQUATION AND BOUNDARY CONDITIONS FOR THE DISTRIBUTION FUNCTION

In this section we will give a brief outline of the derivation of the partial differential equation for the ion-pair distribution function $f(r, \theta)$ and we will also discuss briefly the boundary conditions on f. A background to the theory of dissociation of a weak electrolyte in an applied electric field has been given by Onsager and Fuoss,[8] Onsager[2,7] and Harned and Owen.[9]

Consider an electrolytic solution containing ions of species i and j and let $\mathbf{r}_2$ and $\mathbf{r}_1$ be two position vectors with respect to an arbitrary origin. Then the distribution function f_{ji} is defined as the probability of finding simultaneously an i ion in a unit volume at $\mathbf{r}_2$ and a j ion in a unit volume at $\mathbf{r}_1$. As we will be concerned with the case when an applied electric field acts on the ions, a particular direction in space will be specified and f_{ji} will depend on the direction of $\mathbf{r}_{21} = \mathbf{r}_2 - \mathbf{r}_1$ as well as on its magnitude. The electrolytic solution as a whole can have a nonzero velocity **U**, but for simplicity we will assume that there is no velocity gradient applied to the solution as a whole (as in the theory of viscosity[8]) so that **U** is constant. Thus f_{ji} depends only on the relative position vector $\mathbf{r}_{21}$ and not explicitly on the location $\mathbf{r}_2$ of the i ion; hence $f_{ji} = f_{ji}(\mathbf{r}_{21})$. In the same way we can define the reciprocal function $f_{ij}(\mathbf{r}_{12})$. Clearly $f_{ji}(\mathbf{r}_{21}) = f_{ij}(\mathbf{r}_{12})$.

 0022-2488/78/1906-1340$1.00

The fundamental equation is the continuity equation for ionic motion, which for the case under consideration can be written as

$$\frac{\partial f_{ji}}{\partial t} + \mathrm{div}_2(\mathbf{v} f_{ji}) = 0, \tag{2.1}$$

where $\mathbf{v}$ is the velocity of the i ion relative to the j ion and the subscript 2 denotes differentiation with respect to the components of $\mathbf{r}_2$. For a steady state, which we will be concerned with here, $\partial f_{ji}/\partial t = 0$, and Eq. (2.1) reduces to

$$\mathrm{div}_2(\mathbf{v} f_{ji}) = 0. \tag{2.2}$$

Consider next the velocity $\mathbf{v}$. Denote by $\mathbf{v}_{ji} = \mathbf{v}_{ji}(\mathbf{r}_{21})$ the velocity of an i ion in the neighborhood of a j ion and by $\mathbf{v}_{ij} = \mathbf{v}_{ij}(\mathbf{r}_{12})$ the velocity of a j ion in the neighborhood of an i ion; then $\mathbf{v} = \mathbf{v}_{ji} - \mathbf{v}_{ij}$. The factors which cause the ions to move are (i) the applied electric field $\mathbf{X}$, (ii) interionic forces, (iii) concentration gradients, (iv) the flow of the solution as a whole. Now if the mobility of the i ion is ω_i, then a force $\mathbf{K}_i$ acting on it will produce a velocity $\omega_i \mathbf{K}_i$. The force acting on the i ion due to the applied electric field is $e_i\mathbf{X}$ and this produces a velocity $\omega_i e_i \mathbf{X}$. Denote by $\mathbf{k}_{ji}(\mathbf{r}_{21})$ the force acting on the i ion due to its own "atmosphere" and due to the j ion and its "atmosphere"; this will produce a velocity $\omega_i \mathbf{k}_{ji}$. To determine the diffusion velocity due to a concentration gradient, we will take the diffusion constant[10] of the i ion to be $kT\omega_i$, and so the concentration gradient $\mathrm{grad}_2 f_{ji}$ will produce a current of strength $-kT\omega_i \mathrm{grad}_2 f_{ji}$. Since this current is the product of f_{ji} with the diffusion velocity, it follows that the diffusion velocity is $-kT\omega_i \mathrm{grad}_2(\log f_{ji})$. Thus, if the velocity of the solution as a whole is $\mathbf{U}$ (assumed independent of position), then

$$\mathbf{v}_{ji}(\mathbf{r}_{21}) = \mathbf{U} + \omega_i[e_i\mathbf{X} + \mathbf{k}_{ji}(\mathbf{r}_{21}) - kT\mathrm{grad}_2(\log f_{ji}(\mathbf{r}_{21}))] \tag{2.3}$$

and

$$\mathbf{v}_{ij}(\mathbf{r}_{12}) = \mathbf{U} + \omega_j[e_j\mathbf{X} + \mathbf{k}_{ij}(\mathbf{r}_{12}) - kT\mathrm{grad}_1(\log f_{ij}(\mathbf{r}_{12}))], \tag{2.4}$$

where the subscript 1 denotes differentiation with respect to the components of $\mathbf{r}_1$. Now the interesting and important case of an assemblage of stretched dipoles occurs when the Debye length κ^{-1} of the assemblage is such that

$$q \ll \kappa^{-1}, \tag{2.5}$$

where $\kappa^2 = 4\pi(N_i e_i^2 + N_j e_j^2)/DkT$ and N_i and N_j denote the concentrations of "free" ions. Therefore, we will assume that these concentrations are sufficiently small to ensure that the inequality (2.5) is satisfied. By this assumption the effects of the ionic "atmospheres" are negligible. This is particularly applicable when a strong electric field is present, because the field tends to sweep the "atmosphere" away from the ion. In this case, $\mathbf{k}_{ji}$ is given by the Coulomb potential $e_i e_j/Dr$, where $r = |\mathbf{r}_{21}|$; we have

$$\mathbf{k}_{ji}(\mathbf{r}_{21}) = -\mathrm{grad}_2(e_i e_j/Dr), \tag{2.6}$$

$$\mathbf{k}_{ij}(\mathbf{r}_{12}) = -\mathrm{grad}_1(e_i e_j/Dr) = +\mathrm{grad}_2(e_i e_j/Dr). \tag{2.7}$$

On substituting these expressions into Eqs. (2.3) and (2.4), noting that $f_{ij}(\mathbf{r}_{12}) = f_{ji}(\mathbf{r}_{21})$ and $\mathrm{grad}_1 \equiv -\mathrm{grad}_2$, and subtracting Eq. (2.4) from Eq. (2.3), we obtain

$$\mathbf{v} = \mathbf{v}_{ji} - \mathbf{v}_{ij} = (\omega_i + \omega_j)kT[2\beta\mathbf{n} + \mathrm{grad}_2(2q/r) - \mathrm{grad}_2(\log f_{ji})], \tag{2.8}$$

where q and β are defined by Eqs. (1.2) and (1.3) and

$$\mathbf{n} = \mathrm{sgn}(\omega_i e_i - \omega_j e_j)\mathbf{X}/X, \tag{2.9}$$

where $\mathrm{sgn}(\omega_i e_i - \omega_j e_j) = +1$ if $\omega_i e_i - \omega_j e_j > 0$ and -1 if $\omega_i e_i - \omega_j e_j < 0$. We will assume that $e_i > 0$ and $e_j < 0$, so that $\mathrm{sgn}(\omega_i e_i - \omega_j e_j) = +1$, and $\mathbf{n}$ is a unit vector parallel to the electric field vector $\mathbf{X}$.

Now multiply Eq. (2.8) by f_{ji}, substitute the resulting expression into Eq. (2.2), and suppress the indices i, j and the suffix 2; we find that

$$\nabla^2 f = 2\epsilon\,\mathrm{div}(f\mathbf{n}) + \mathrm{div}\left(f\,\mathrm{grad}\,\frac{1}{r}\right), \tag{2.10}$$

where r has been made nondimensional on division by the characteristic length $2q$ and $\epsilon = 2\beta q$. As $\mathrm{div}\,\mathbf{n} = 0$, as $\mathbf{n}$ is a constant vector, and $\nabla^2(1/r) = 0$, Eq. (2.10) can be rewritten as

$$\nabla^2 f = \left(2\epsilon\mathbf{n} + \mathrm{grad}\,\frac{1}{r}\right)\cdot \mathrm{grad} f. \tag{2.11}$$

Finally we rewrite Eq. (2.11) in terms of spherical polar coordinates (r, θ, ϕ), where θ is the angle the ion pair makes with the electric field direction $\mathbf{n}$. Due to symmetry about $\mathbf{n}$, there is no dependence of f on ϕ; Eq. (2.11) becomes

$$r^2\frac{\partial^2 f}{\partial r^2} + (1 + 2r - 2\epsilon r^2\cos\theta)\frac{\partial f}{\partial r} + \frac{\partial^2 f}{\partial \theta^2} + (\cot\theta + 2\epsilon r\sin\theta)\frac{\partial f}{\partial \theta} = 0. \tag{2.12}$$

This is the required partial differential equation for $f(r, \theta)$.

The distribution function is the sum of two parts, both of which satisfy Eq. (2.12): the part $n_i n_j$ describing the "free" ions and corresponding to complete dissociation,[2] and the part describing associated ions. We will only be concerned here with the later part. This satisfies the boundary conditions

$$\lim_{r\to\infty} f(r, \theta) = 0 \tag{2.13}$$

corresponding to the state of complete dissociation at infinity, and

$$\lim_{r\to 0} e^{\alpha 1/r} f(r, \theta) = 1. \tag{2.14}$$

The boundary condition (2.14) can be justified as follows. Due to the local equilibrium[11] in the region $r \simeq a$, where a is the distance of closest approach of the ions in an ion pair, the total distribution function must correspond to a Maxwell–Boltzmann distribution in the combined Coulomb and external field for this r. Thus for $r \simeq a$, we must have

$$f(r, \theta) \sim \exp\left(\frac{1}{r} + 2\epsilon r\cos\theta\right). \tag{2.15}$$

For the ideal case $a = 0$, this condition reduces to the boundary condition (2.14). The limit in (2.14) is unity

due to the normalization condition imposed on f in Sec. 1.

To summarize, the mathematical problem is to solve the partial differential equation (2.12) for $f(r,\theta)$ subject to the boundary conditions (2.13) and (2.14).

3. SUMMARY OF RESULTS

In this section we summarize briefly previously obtained[6] results which we require to determine the ion-pair distribution function.

In solving the related problem of Wien dissociation in a weak electrolyte, we used a perturbation technique, the perturbation parameter being $\epsilon = 2\beta q$. Now, if $\cos\theta \neq 1$, $f(r,\theta)$ does not possess a perturbation expansion in powers of ϵ which is valid for large r. For it can be shown[6] by considering the asymptotic solution of Eq. (2.12) that, for large r,

$$f(r,\theta) \sim \frac{1}{r}\exp[\epsilon r(\cos\theta - 1)]. \tag{3.1}$$

If we expand the right-hand side of (3.1) in powers of ϵ, we see that the individual terms of this expansion do not satisfy the boundary condition (2.13) for large r.

Further, from the boundary condition (2.14), it follows that $f(r,\theta)$ diverges exponentially to infinity as r tends to zero. We therefore introduced the function $h(r,\theta)$ defined by

$$f(r,\theta) = h(r,\theta)\exp[1/r + \epsilon r(\cos\theta - 1)]; \tag{3.2}$$

unlike $f(r,\theta)$, the function $h(r,\theta)$ has a perturbation expansion in ϵ which is valid for $0 \leq r \leq \infty$.

It proves convenient to define $x = \cos\theta$. In our previous paper,[6] we let $u = 1/r$, but no advantage is gained by making that change of variable in the present discussion. Our method of solution consisted in taking the Legendre transform of the partial differential equation for $h(r,x)$ by operating on it with $\int_{-1}^{1} dx P_n(x)$, where $P_n(x)$ is the Legendre polynomial of degree n; this replaced the partial differential equation by an infinite system of ordinary differential equations. We expanded $h(r,x)$ in the form

$$h(r,x) = \sum_{m=0}^{\infty} \epsilon^m h^{(m)}(r,x), \tag{3.3}$$

and defined the Legendre transform

$$H_n^{(m)}(r) = \int_{-1}^{1} P_n(x) h^{(m)}(r,x)\,dx. \tag{3.4}$$

The ordinary differential equation which we obtained in this way for $H_n^{(m)}(r)$ was solved by looking for a solution in which $h^{(m)}(r,x)$ is separable in r and x. It is not necessary for the problem to admit such a solution for the perturbation technique to work, but it simplifies the analysis if it does. Suppose

$$h^{(m)}(r,x) = F^{(m)}(r) Z^{(m)}(x) \tag{3.5}$$

and expand $Z^{(m)}(x)$:

$$Z^{(m)}(x) = \sum_{s=0}^{\infty} c_s^{(m)} P_s(x), \tag{3.6}$$

where $c_s^{(m)}$ are constants. Then it follows from the orthogonality property[12] of Legendre polynomials that

$$H_n^{(m)}(r) = \frac{2}{(2n+1)} c_n^{(m)} F^{(m)}(r). \tag{3.7}$$

It is considerably easier to solve for $F^{(m)}(r)$ and the constants $c_n^{(m)}$ than for $H_n^{(m)}(r)$. Using induction, we found that

$$F^{(m)}(r) = \frac{2^m r^m}{(m+1)!}\left(1 - \sum_{p=0}^{m} \frac{e^{-1/r}}{p!\, r^p}\right) \tag{3.8}$$

and

$$c_n^{(m)} = \begin{cases} \dfrac{(2n+1)m!\,(m+1)!}{(m-n)!\,(m+n+1)!}, & 0 \leq n \leq m, \qquad (3.9a) \\ 0, & n \geq m+1. \qquad (3.9b) \end{cases}$$

In the calculation of $K(X)/K(0)$ in our previous paper, all that was required was a knowledge of $F^{(m)}(r)$ and $c_n^{(m)}$. We will now show how the above results can be used to determine the ion-pair distribution function itself.

4. SOLUTION FOR $h(r,x)$

It follows from Eqs. (3.6) and (3.9) that

$$Z^{(m)}(x) = \sum_{n=0}^{m} \frac{(2n+1)m!\,(m+1)!}{(m-n)!\,(m+n+1)!} P_n(x), \tag{4.1}$$

and hence, using Eq. (3.5) and (3.8), we have

$$h^{(m)}(r,x) = 2^m m!\, r^m \left(1 - \sum_{p=0}^{m} \frac{e^{-1/r}}{p!\, r^p}\right) \times \sum_{n=0}^{m} \frac{(2n+1)P_n(x)}{(m-n)!\,(m+n+1)!}. \tag{4.2}$$

The summation involving Legendre polynomials in Eq. (4.2) can be evaluated using induction. We will make the following induction assumption: We will assume that for some integer $m \geq 0$,

$$\sum_{n=0}^{m} \frac{(2n+1)P_n(x)}{(m-n)!\,(m+n+1)!} = \frac{(1+x)^m}{2^m (m!)^2}. \tag{4.3}$$

Clearly the induction assumption is satisfied for $m = 0$, for then both sides equal unity. To show the result is valid for $m+1$, assuming it to be true for some integer $m \geq 0$, multiply Eq. (4.3) by $(1+x)$; then

$$\sum_{n=0}^{m} \frac{(2n+1)(1+x)P_n(x)}{(m-n)!\,(m+n+1)!} = \frac{(1+x)^{m+1}}{2^m (m!)^2}. \tag{4.4}$$

But we can replace $xP_n(x)$ in (4.4) by[12]

$$xP_n(x) = \frac{(n+1)}{(2n+1)} P_{n+1}(x) + \frac{n}{(2n+1)} P_{n-1}(x); \tag{4.5}$$

we also have

$$\sum_{n=0}^{m} \frac{(2n+1)P_n(x)}{(m-n)!\,(m+n+1)!} = \sum_{n=0}^{m+1} \frac{(2n+1)(m+1-n)P_n(x)}{(m+1-n)!\,(m+n+1)!}, \tag{4.6}$$

$$\sum_{n=0}^{m} \frac{(n+1)P_{n+1}(x)}{(m-n)!\,(m+n+1)!} = \sum_{n=0}^{m+1} \frac{nP_n(x)}{(m+1-n)!\,(m+n)!}, \tag{4.7}$$

$$\sum_{n=1}^{m} \frac{nP_{n-1}(x)}{(m-n)!\,(m+n+1)!} = \sum_{n=0}^{m+1} \frac{(n+1)(m-n)(m+1-n)P_n(x)}{(m+1-n)!\,(m+n+2)!}. \tag{4.8}$$

Thus adding together Eqs. (4.6), (4.7), and (4.8), we find, after some further algebra, that

$$\sum_{n=0}^{m} \frac{(2n+1)(1+x)P_n(x)}{(m-n)!(m+n+1)!}$$

$$= 2(m+1)^2 \sum_{n=0}^{m+1} \frac{(2n+1)P_n(x)}{(m+1-n)!(m+n+2)!} \cdot \qquad ((4.9)$$

Substituting this result into Eq. (4.4), we obtain

$$\sum_{n=0}^{m+1} \frac{(2n+1)P_n(x)}{(m+1-n)!(m+1+n+1)!} = \frac{(1+x)^{m+1}}{2^{m+1}[(m+1)!]^2}. \qquad (4.10)$$

Thus, by induction, identity (4.3) is valid for all integers $m \geq 0$.

To complete the solution for $h(r,x)$, we see from Eqs. (4.2) and (4.3) that

$$h^{(m)}(r,x) = \frac{(1+x)^m r^m}{m!}\left(1 - \sum_{p=0}^{m} \frac{e^{-1/r}}{p!\, r^p}\right), \qquad (4.11)$$

and hence from Eq. (3.3), we obtain

$$h(r,x) = \sum_{m=0}^{\infty} \frac{\epsilon^m (1+x)^m r^m}{m!}\left(1 - \sum_{p=0}^{m} \frac{e^{-1/r}}{p!\, r^p}\right). \qquad (4.12)$$

5. ION-PAIR DISTRIBUTION FUNCTION AND COMPARISON WITH ONSAGER'S RESULT

It follows from Eqs. (3.2) and (4.12) that

$$f(r,x) = \exp\left(\frac{1}{r} + \epsilon r(x-1)\right)$$

$$\times \sum_{m=0}^{\infty} \frac{\epsilon^m (1+x)^m r^m}{m!}\left(1 - \sum_{p=0}^{m} \frac{e^{-1/r}}{p!\, r^p}\right). \qquad (5.1)$$

On expanding the exponential in Eq. (5.1), we obtain

$$f(r,x) = \sum_{m=0}^{\infty} \sum_{s=0}^{\infty} \sum_{p=m+1}^{\infty} \frac{\epsilon^{m+s} r^{m+s-p}}{m!\, s!\, p!} (x+1)^m (x-1)^s, \qquad (5.2)$$

and if we write

$$f(r,x) = \sum_{n=0}^{\infty} \epsilon^n f^{(n)}(r,x), \qquad (5.3)$$

then we find that

$$f^{(n)}(r,x) = \sum_{m=0}^{\infty} \sum_{p=m+1}^{\infty} \frac{r^{n-p}(x+1)^m (x-1)^{n-m}}{m!\,(n-m)!\, p!}. \qquad (5.4)$$

Expression (5.1) is valid for all $0 \leq r \leq \infty$; in particular it satisfies the boundary conditions (2.13) and (2.14). As stated previously, (5.3) is not a valid perturbation expansion in ϵ when r is large. For, if $x \neq 1$ and $n \geq 2$, $f^{(n)} \to \infty$ like r^{n-1} as $r \to \infty$ and therefore does not satisfy the boundary condition (2.13) for $r = \infty$. There are some applications, however, in which the perturbation expansion of f in powers of ϵ need not be valid for all r. For example, $K(X)/K(0)$ depends on $f(r,\theta)$, but is independent of r. Hence we can use (5.3) to evaluate $K(X)/K(0)$ as a perturbation expansion in powers of ϵ, by choosing an r, say of order unity, for which expansion (5.3) is valid.

Finally, to compare Onsager's solution (1.1) with Eq. (5.1), we first note that

$$J_0(y) = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m!)^2}\left(\frac{y}{2}\right)^{2m}; \qquad (5.5)$$

thus

$$\int_0^1 J_0\left[(-8\epsilon s)^{1/2} \cos\frac{\theta}{2}\right] e^{-s/r} ds$$

$$= \sum_{m=0}^{\infty} \frac{\epsilon^m (1+x)^m}{(m!)^2} \int_0^1 s^m e^{-s/r} ds. \qquad (5.6)$$

But, using induction, or otherwise, it can be shown that

$$\int_0^1 s^m e^{-s/r} ds = m!\, r^{m+1}\left(1 - \sum_{p=0}^{m} \frac{e^{-1/r}}{p!\, r^p}\right). \qquad (5.7)$$

On substituting Eq. (5.7) into Eq. (5.6) and then the resulting expression into Eq. (1.1), we see that Onsager's solution (1.1) and our solution (5.1) are the same.

6. CONCLUDING REMARKS

We have given the remaining details in the derivation of the distribution function of an assemblage of ion pairs where the main interaction is electrostatic. Such a result has applicability in the theory of weak electrolytes and in theories of the propagation of the nerve impulse.[13,14] Generalization of this result to nonisotropic media and to the case of ion-exchange resins and soluble polyelectrolytes may also be useful.

In the above analysis, it turned out that, to obtain $h^{(m)}(r,x)$, it was not necessary to use the formula for the inverse Legendre transform. This was because the problem admitted a separable solution of the form (3.5), and hence we merely had to evaluate the summation (3.6). For similar problems in this and related fields, it is possible that a separable solution will not exist. In such cases $h^{(m)}(r,x)$ can be obtained by evaluating the inverse transform[15]

$$h^{(m)}(r,x) = \tfrac{1}{2}\sum_{n=0}^{\infty}(2n+1)H_n^{(m)}(r)P_n(x). \qquad (6.1)$$

It is not difficult to check that for the problem considered in this paper, Eq. (6.1) leads to the same result.

Identity (4.3) is of interest in its own right, as it provides an expansion of $(1+x)^m$ in Legendre polynomials. By giving x particular values, it can be used to evaluate certain summations. For example, on setting $x = +1$ and $x = -1$ and noting that[12] $P_n(1) = 1$ and $P_n(-1) = (-1)^n$, Eq. (4.3) yields respectively

$$\sum_{n=0}^{m} \frac{(2n+1)}{(m-n)!\,(m+n+1)!} = \frac{1}{(m!)^2}, \qquad (6.2)$$

$$\sum_{n=0}^{m} \frac{(-1)^n(2n+1)}{(m-n)!\,(m+n+1)!} = 0. \qquad (6.3)$$

Similar summations can be evaluated with the aid of the results[12] $P_n'(1) = \frac{1}{2}n(n+1)$ and $P_n'(-1) = (-1)^n \frac{1}{2}n(n+1)$, where the prime denotes differentiation with respect to x.

[1] F.W. Sears, *An Introduction to Thermodynamics, the Kinetic Theory of Gases, and Statistical Mechanics* (Addison-Wesley, London, 1959), 2nd ed., p. 309.
[2] L. Onsager, J. Chem. Phys. 2, 599 (1934).
[3] This term is embodied in the mathematics by condition (2.5): since κq is a measure of the interionic forces, condition (2.5) implies that these forces are small for our problem,

which is, of course, so because we suppose association of ions into "bound" pairs is caused mainly by the Coulombic force between the ions of a pair.

[4]At any given instant there is a unique value of θ associated with each dipole, but this value of θ changes with time because the dipoles tend to align themselves with the electric field. In fact by expressing Eq. (2.8) in spherical polar coordinates, it can be shown that for $\beta \neq 0$, we can have $\dot{\theta} = 0$ only if $\theta = 0$ or π, in which case the dipole is aligned parallel, or antiparallel, with the field.

[5]N. Bjerrum, Kgl. Danske Vid. Selsk, Math. Fys. Medd. **7**, 9 (1926).

[6]D.P. Mason and D.K. McIlroy, Proc. Roy. Soc. Lond. A **359** (1978), in press.

[7]L. Onsager "Solutions of the Mathieu equation of period 4π," thesis (Yale University, 1935).

[8]L. Onsager, R.M. Fuoss, J. Phys. Chem. 36, 2689 (1932).

[9]H.S. Harned and B.B. Owen, *Physical Chemistry of Electrolytic Solutions* (Reinhold, New York, 1958), 3rd ed., Chap. 2.

[10]R.C. Tolman, *Statistical Mechanics* (Reinhold, New York, 1927), 1st ed., pp. 229—31.

[11]Local equilibrium in the region $r \simeq a$ is not strictly true if $e_i + e_j \neq 0$, but the error of nevertheless assuming equilibrium in this case is small, being of the order of that already made in neglecting the hydrodynamic interaction of the ions.

[12]W.W. Bell, *Special Functions for Scientists and Engineers* (Van Nostrand, London, 1968), Chap. 3.

[13]D.K. McIlroy, Math. Biosci. **7**, 313 (1970).

[14]D.K. McIlroy, Math. Biosci. **8**, 109 (1970).

[15]C.J. Tranter, *Integral Transforms in Mathematical Physics* (Methuen, London, 1966), 3rd ed., p. 96.

Linear response theory revisited. I. The many-body van Hove limit

K. M. van Vliet

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada
(Received 19 October 1977)

A critical discussion of linear response theory is given. It is argued that in the formalism as it stands no dissipation is manifest. A physical reinterpretation for the case of a system in weak interaction with a reservoir is given. Mathematically this means that the van Hove limit, as well as the large system limit, is applied to the time-dependent Heisenberg operators of the Kubo formalism. The reduced operators can be put in a very compact form, viz., $B_a^R(t) = [\exp(-\Lambda_d t)]B_a$, where B_a is a Schrödinger operator and Λ_d is the Liouville space superoperator corresponding to the transition operator of the master equation. In this form the relaxation character of the transport expressions, and the approach to equilibrium is at once evident. New expressions for the generalized susceptibility and conductivity in this limit are presented. Also, the Onsager relations and other symmetry properties are confirmed.

1. INTRODUCTION

Some twenty years ago Kubo published his now classic papers on linear response theory.[1,2] In these papers, he developed a general many-body formalism for expressing transport coefficients in terms of correlation functions of the transport quantities involved. His expressions generalized earlier results by Green,[3] Kirkwood,[4] and others. Solving von Neumann's equations, the response $\langle \Delta B(t) \rangle$ of some operator B due to a field or other ponderomotive force is found to be expressible in terms of a response function, relaxation function, or correlation function of the spontaneous fluctuations $\langle \Delta B(t) \Delta B(0) \rangle$ in the system. The latter can in principle be determined as an ensemble average in a canonical ensemble, for which the density operator is known if the Hamiltonian is known. In this manner very general microscopic expressions can be given for the transport coefficients.

The correlation expressions can also be Fourier analyzed; the results can then be shown to be equivalent with the fluctuation–dissipation theorem which gives the connection between the dissipative generalized conductance of a process and the spectral density of the spontaneous fluctuations of the system. Thus, *the Kubo relations do in the time domain what the fluctuation-dissipation theorem does in the frequency domain.* The fluctuation–dissipation theorem in its general form is due to Callen and Welton[5] and Greene and Callen[6]; dealing only with electrical conduction, it goes back to Nyquist,[7] and even to Einstein and de Haas-Lorentz.[8]

Yet, the Kubo relations have a special appeal since they deal directly with the microscopic quantum mechanical motion of a process. Moreover, the derivation is by many believed to be exact, except for the linearization in the applied field, which, *a prima vista* looks no worse than the similar procedure followed in the usual perturbation solution of the Boltzmann equation. Strong interactions can in principle be included in the computations of the transport coefficients, cf. Verboven[9] and Fujita and Abe.[10]

However, it is precisely the generality of the expressions, which has invited criticism. Van Kampen[11] has emphasized that there is a vast difference between microscopic linearity and macroscopic linearity of the responses. Also, for applications it is usually found that Kubo's expressions are too general, so that somewhere in the application a randomness assumption must be made. In this respect the various attempts to compare Kubo's formulas with Boltzmann-type results are illuminating, cf. Refs. 9, 10, 12–15. Kubo's formulas are then first reduced to the corresponding one-particle expressions, in which subsequently various randomizing effects are incorporated, see especially Chester and Thellung.[12]

In this paper we will set forth a new perspective of linear response theory. We believe that the formalism in its generality can probably not be justified; however, the framework can be maintained, but is in need of a reinterpretation. In particular, we introduce a randomness assumption on the many-body level, by considering weak interaction with a reservoir. It turns out that van Hove's formalism[16] developed in connection with his derivation of the Pauli master equation, is all but tailored for this purpose. We will show that there is a drastic change in the time dependence of the operators (reduced Heisenberg operators) after the application of the van Hove weak coupling limit. The ensuing transport formulas now clearly show the effects of dissipation and relaxation. The results can, if one so wishes, be applied to one-particle cases. No new stosszahlansatz is then necessary, the reduction going similarly as the derivation of the Boltzmann equation from the master equation by van Hove;[17] this is planned for future work.

In this paper Kubo's expression, rather than the master equation, is employed as the point of departure; nevertheless, the connection with the master equation formalism is very close. Indeed, in a subsequent paper,[18] we will obtain almost the same results, by not using von Neumann's equation, which was the basis of Kubo's theory, but a new inhomogeneous master equation (i.e., containing streaming terms) as the point of departure. Thus, *our assertion is that linear response theory should be put on a master equation level, thereby acquiring stochastic content, prior to its application for the calculation of transport coefficients.*

Finally, a note on extensions. In case the interactions with the reservoir are strong, a generalized master equation approach, using projection operator and resol-

 0022-2488/78/1906-1345$1.00

vent techniques appears promising. However, such results do not show the clear irreversible behavior and stochastic character of the present theory. In terms of physical understanding, these extensions are therefore much less fundamental.

This article is divided into three main parts. In part A we consider the original Kubo theory adapted to our present needs and insights. In Part B we voice various criticisms on the formalism and we undertake to give a reinterpretation; the reduction of the Heisenberg operators is carried out, and the new, dissipative behavior and the approach to equilibrium are explicitly shown. In Part C, finally, revised response formulas are established in two forms, indicated as the interaction form and the Schrödinger form. In the latter form the stochastic nature of the expressions is most clearly revealed. We also show that the new formalism, being a synthesis of Kubo's and van Hove's pioneering work, gives a fully microscopic basis for Onsager's relations. In the appendices we retrace and extend van Hove's original paper in support of the derivations of Part B.

A. ELEMENTS OF KUBO THEORY

2. THE RESPONSE FUNCTION AND THE RELAXATION FUNCTION

Prior to our reinterpretation, we will summarize the main aspects of the formalism as it stands, cf. Kubo,[1,2,19] Mazo,[20] and Montroll.[21]

Kubo considers systems with a Hamiltonian

$$H_{\text{total}} = H - AF(t), \tag{2.1}$$

where H is the Hamiltonian of the system proper and $AF(t)$ is the coupling with the external field; A is an operator corresponding to some observable $\mathcal{A}$ and $F(t)$ a (complex) time function (c-number). In certain situations **A** and **F** are vectors. For example, for a system in an electric field we have $\mathbf{A}\cdot\mathbf{F}(t) = \sum_i q_i \mathbf{r}_i \cdot \mathbf{E}(t)$ in which $\mathbf{r}_i$ is the position of charge q_i and $\mathbf{E}(t)$ is the time-dependent field. Another example is that of magnetic dipoles in a magnetic field, $\mathbf{A}\cdot\mathbf{F}(t) = \sum_i \mu_i \cdot \mathbf{B}(t)$.

In Kubo's papers, as well as other treatments we know of, the field is assumed to be turned on at $t = -\infty$. The response of an observable $\mathcal{B}$, represented by an operator B, at time t is then sought for. However, these treatments seem to be unaware of standard response theory as developed for networks in electrical engineering, see, e.g., van Valkenburg.[22] It is common practice to turn on the disturbance at $t = 0$. The transient response which follows is best approached with Laplace transforms.

Thus, following these ideas, we seek the solution of the von Neumann equation for the Hamiltonian (2.1)

$$\frac{\partial \rho}{\partial t} + \left(\frac{i}{\hbar}\right)[H, \rho] = \left(\frac{i}{\hbar}\right) u(t) F(t) [A, \rho], \tag{2.2}$$

where $\rho(t)$ is the density operator and $u(t)$ is the unit step function or Heaviside function, being zero for $t < 0$ and unity for $t > 0$. A formal equivalent of (2.2) is the Volterra integral equation

$$\rho = \rho_0(t) + \left(\frac{i}{\hbar}\right) u(t) \int_0^t \exp[-iH(t-t')/\hbar] \times F(t')[A, \rho(t')] \exp[iH(t-t')/\hbar]\, dt', \tag{2.3}$$

where $\rho_0(t)$ is the solution of the homogeneous equation, i.e., when the rhs of (2.2) is zero. Equation (2.3) is solved by iterating once. Thus, the linearized solution is found by substituting $\rho(t) \approx \rho_0(t)$ in the right-hand side.

As to $\rho_0(t)$, it is assumed that at $t = 0-$ the system has thermalized a long time so that $\rho_0(t)$ is the equilibrium density operator ρ_{eq}. A problem with the (time-reversible) von Neumann equation is that there is no approach to equilibrium. Basically, however, we believe that the von Neumann equation applies to a closed system, thus indicating the need for a microcanonical ensemble (a closed system in the terminology of Gibbs[23] may still include energies which are mutual to the system and external bodies, such as energies due to external fields). The fact that Kubo instead employs a canonical ensemble for the unperturbed density operator, can only be justified in the thermodynamic limit; see also Sec. 4.2.

With these cautionings, we then find for the response of B caused by the switched on field for $t > 0$,

$$\begin{aligned}\langle \Delta B(t) \rangle &= \mathrm{Tr}(\rho(t)B) - \mathrm{Tr}(\rho_{\text{eq}}B) \\ &= (i/\hbar)\,\mathrm{Tr} \int_0^t dt'\, B \exp[-iH(t-t')/\hbar] \\ &\quad \times F(t')[A, \rho_{\text{eq}}] \exp[iH(t-t')/\hbar],\end{aligned} \tag{2.4}$$

with

$$\rho_{\text{eq}} = \exp(-\beta H)/\mathrm{Tr}[\exp(-\beta H)], \quad \beta = 1/kT. \tag{2.5}$$

We now use the fact that the trace is invariant under cyclic permutation, $\mathrm{Tr}ABC\cdots = \mathrm{Tr}BC\cdots A$, etc. We then find

$$\begin{aligned}\langle \Delta B(t) \rangle &= (i/\hbar)\mathrm{Tr} \int_0^t d\tau\, B(t-\tau)[A, \rho_{\text{eq}}] F(\tau) \\ &= (1/\hbar i)\mathrm{Tr} \int_0^t d\tau [A, B(t-\tau)] \rho_{\text{eq}} F(\tau),\end{aligned} \tag{2.6}$$

where $B(t)$ is the Heisenberg operator[24]

$$B(t) = \exp(iHt/\hbar)\, B \exp(-iHt/\hbar). \tag{2.7}$$

Kubo refers to $B(t)$ as the "natural motion" of the observable $\mathcal{B}$.

For later reference we also indicate the Liouville solution. The lhs of (2.2) can also be written as $\partial\rho/\partial t + i\mathcal{L}\rho$, where $\mathcal{L}$ is the Liouville operator, defined by

$$i\mathcal{L}\, C = \{H, C\} \text{ (classical) or } = (1/\hbar)[H, C] \text{ (quantum mechanical)}, \tag{2.8}$$

where C is an arbitrary operator. For the quantum mechanical case one finds

$$\langle \Delta B(t) \rangle = (i/\hbar)\,\mathrm{Tr} \int_0^t dt'\, B \exp[-i(t-t')\mathcal{L}] \times [A, \rho_{\text{eq}}] F(t'). \tag{2.9}$$

Comparison with (2.4) shows the operator identity

$$\exp(-i\mathcal{L}t)C = \exp(-iHt/\hbar)C\exp(iHt/\hbar), \tag{2.10}$$

so that in particular we have

$$B(t) = \exp(i\mathcal{L}t)B. \tag{2.11}$$

We note that $\mathcal{L}$ is a superoperator, acting on operators of the Liouville space $H\otimes\bar{H}$, with H being the Hilbert space spanned by the Hamiltonian and with $\bar{H}$ being the dual space. The properties of $\mathcal{L}$ in the Liouville space have been well studied by Fano.[25] Since $\mathcal{L}$ is Hermitian, $\exp(i\mathcal{L}t)$ is unitary. The eigenvalues of $\mathcal{L}$ are the eigenfrequencies

$$\omega_{rs} = (\mathcal{E}_r - \mathcal{E}_s)/\hbar = \pm|\omega_r - \omega_s|, \tag{2.12}$$

where $\mathcal{E}_k$ are the energies of H. The eigenvalues of $\exp(i\mathcal{L}t)$ are imaginary with modulus one, so that $B(t)$ has conserved norm. Kubo's "natural motion" thus constitutes a rotation in the Liouville space.

Returning to (2.6), we introduce the response function

$$\phi_{BA}(t) = (1/\hbar i)\,\mathrm{Tr}\{[A, B(t)]\rho_{eq}\}, \tag{2.13}$$

so we have the convolution integral

$$\langle\Delta B(t)\rangle = \int_0^t d\tau\,\phi_{BA}(t-\tau)F(\tau). \tag{2.14}$$

Clearly ϕ_{BA} is a memory function. With $\hat{F}(s)$, $b(s)$, and $\chi_{BA}(s)$ being the Laplace transforms of $F(t)$, $\Delta B(t)$, and $\phi_{BA}(t)$, respectively, (2.14) yields

$$b(s) = \chi_{BA}(s)\hat{F}(s). \tag{2.15}$$

Thus, for the *generalized susceptibility* χ_{BA}, we have

$$\chi_{BA}(s) = \int_0^\infty dt\exp(-st)\phi_{BA}(t)$$
$$= (1/\hbar i)\int_0^\infty dt\exp(-st)\,\mathrm{Tr}\{[A, B(t)]\rho_{eq}\}. \tag{2.16}$$

We note that we did not do much physics, since we said nothing about the processes which caused or sustained the response; we found, however, a general result in terms of a commutator of a correlation expression.

Kubo also introduces the relaxation function. If a constant perturbation acts from $t=-\infty$ to $t=0$, at which moment the perturbation is switched off, we have $F(t) = F_0[1-u(t)]$; the response is given by

$$\langle\Delta B(t)\rangle = \int_{-\infty}^t d\tau\,\phi_{BA}(t-\tau)F_0[1-u(\tau)]$$
$$= F_0\int_{-\infty}^0 \phi_{BA}(t-t')\,dt' = F_0\int_t^\infty \phi_{BA}(\tau)\,d\tau$$
$$\equiv F_0\Psi_{BA}(t), \tag{2.17}$$

where we set $\tau = t - t'$. The function Ψ_{BA} is the relaxation function. With (2.13) one finds

$$\Psi_{BA}(t) = \int_t^\infty dt'\,\phi_{BA}(t') = \frac{1}{\hbar i}\int_t^\infty dt'\,\mathrm{Tr}\{[A, B(t')]\rho_{eq}\}, \tag{2.18}$$

or also

$$\frac{d}{dt}\Psi_{BA}(t) = -\frac{1}{\hbar i}\mathrm{Tr}\{[A, B(t)]\rho_{eq}\}. \tag{2.19}$$

Often one seeks the response of a flux or current $d\beta/dt$ represented[26] by an operator $\dot{B}$. Then, analogous to (2.14)

$$\langle\Delta\dot{B}(t)\rangle = \int_0^t d\tau\,\phi_{\dot{B}A}(t-\tau)F(\tau), \tag{2.20}$$

with the response function

$$\phi_{\dot{B}A}(t) = \frac{1}{\hbar i}\mathrm{Tr}\{[A, \dot{B}(t)]\rho_{eq}\}. \tag{2.21}$$

For this case it is more physical (though not necessary) to describe the response by a *generalized conductance* L_{BA}, rather than by χ. Thus, making a Laplace transformation of (2.20), we have

$$\dot{b}(s) = L_{BA}(s)\hat{F}(s), \tag{2.21a}$$

with

$$L_{BA}(s) = \int_0^\infty dt\exp(-st)\phi_{\dot{B}A}(t)$$
$$= (1/\hbar i)\int_0^\infty dt\exp(-st)\,\mathrm{Tr}\{[A, \dot{B}(t)]\rho_{eq}\}. \tag{2.22}$$

We have the following identical functions,

$$\phi_{\dot{B}A} = \dot{\phi}_{BA} = -\ddot{\Psi}_{BA}. \tag{2.23}$$

These identities are evident. We can further transform as follows,

$$\ddot{\Psi}_{BA} = \frac{d}{dt}\dot{\Psi}_{BA} = -\frac{1}{\hbar i}\mathrm{Tr}\{[A, \dot{B}(t)]\rho_{eq}\}$$
$$= -\frac{1}{\hbar i}\mathrm{Tr}\{[A(-t), \dot{B}]\rho_{eq}\} = \frac{1}{\hbar i}\int_t^\infty dt'\frac{d}{dt'}\mathrm{Tr}\{[A(-t), \dot{B}]\rho_{eq}\}$$
$$= -\frac{1}{\hbar i}\int_t^\infty dt'\,\mathrm{Tr}\{[\dot{A}, \dot{B}(t)]\rho_{eq}\}. \tag{2.24}$$

We introduce the function

$$\phi_{\dot{B}\dot{A}}(t) \equiv (1/\hbar i)\,\mathrm{Tr}\{[\dot{A}, \dot{B}(t)]\rho_{eq}\}. \tag{2.25}$$

This definition is motivated by (2.13). Notice that this is a "would-be" function: It does not pertain to the process itself, but would arise if there were an external Hamiltonian $\dot{A}F(t)$. In the various transitions of (2.24) we used the following two properties:

(i) *Stationarity*: As for any two-point correlation, we have for two operators C and D

$$\mathrm{Tr}\{\rho_{eq}C(t_1)D(t_2)\} = \mathrm{Tr}\{\rho_{eq}C(t_1-t)D(t_2-t)\}. \tag{2.26}$$

The proof follows immediately from the definition of the Heisenberg operators (2.7) and from cyclic permutivity. In particular (2.26) indicates

$$\mathrm{Tr}\{\rho_{eq}[A(0), \dot{B}(t)]\} = \mathrm{Tr}\{\rho_{eq}[A(-t), \dot{B}(0)]\}. \tag{2.27}$$

(ii) A form of the *mixing property*: It is assumed that

$$\lim_{|t_1-t_2|\to\infty}\mathrm{Tr}\{\rho_{eq}C(t_1)D(t_2)\} = \langle C\rangle\langle D\rangle; \tag{2.28}$$

in particular this indicates that the limit of a commutator vanishes,

$$\lim_{t\to\infty}\mathrm{Tr}\{\rho_{eq}[A(-t), \dot{B}(0)]\} = 0. \tag{2.29}$$

This mixing property is much discussed by Kubo (Ref. 1, p. 577). In its generality this property has not been proven as yet. A proof for harmonic crystals is given by Lanford and Lebowitz.[27] We come back to this property in Sec. 8.

We substitute the last line of (2.24) into (2.23). We then obtain for the generalized conductance

$$\begin{aligned} L_{BA}(s) &= \frac{1}{\hbar i}\int_0^\infty dt \int_t^\infty dt' \exp(-st)\mathrm{Tr}\{[\dot{A}, \dot{B}(t')]\rho_{eq}\} \\ &= \frac{1}{\hbar i}\int_0^\infty dt' \int_0^{t'} dt \exp(-st)\mathrm{Tr}\{[\dot{A}, \dot{B}(t')]\rho_{eq}\} \\ &= -\int_0^\infty dt \frac{\exp(-st)-1}{s\hbar i}\mathrm{Tr}\{[\dot{A}, \dot{B}(t)]\rho_{eq}\} \\ &= -\int_0^\infty dt \frac{\exp(-st)-1}{s}\phi_{\dot{B}\dot{A}}(t), \end{aligned} \tag{2.30}$$

where we changed the order of integrations in the second step.

This result applies to the electrical conductivity, in which case $A = \sum_i q_i \mathbf{r}_i$, $\dot{A} = \sum_i q_i \mathbf{v}_i = \mathbf{J}$. The current density for a crystal of volume V_0 containing N electrons is

$$\mathbf{J} = \frac{1}{N}\sum_i q_i \mathbf{v}_i \frac{N}{V_0} = \frac{\dot{\mathbf{A}}}{V_0}. \tag{2.31}$$

Since L is defined by $\mathbf{J}V_0 = \mathsf{L}_{\mathbf{AA}} \cdot \mathbf{E}$ (where L is a tensor) while also $\mathbf{J} = \boldsymbol{\sigma} \cdot \mathbf{E}$, we see that $\boldsymbol{\sigma} = \mathsf{L}_{\mathbf{AA}}/V_0$. Hence from (2.30),

$$\sigma_{\mu\nu}(s) = -V_0 \int_0^\infty dt \frac{\exp(-st)-1}{s\hbar i}\mathrm{Tr}\{[J_\nu, J_\mu(t)]\rho_{eq}\}, \tag{2.32}$$

where Greek subscripts refer to tensor components.

3. THE FREQUENCY DOMAIN; VARIOUS FORMS

We give a variety of expressions for χ and L. We will emphasize the generic relationship of the various results; the reason for giving several is that some of these expressions are suitable for carry-over in the new results (Sec. 9), whereas others are necessary for the statement of the fluctuation–dissipation theorem.

3.1. The commutator form

For sinusoidal excitation we have $F(t) = u(t)F_{0,\omega} \times \exp(i\omega t)$. For $t \to \infty$ the response is also sinusoidal, $\langle \Delta B(t)\rangle = B_{0,\omega}\exp(i\omega t)$, with

$$B_{0,\omega} = \chi(i\omega)F_{0,\omega}, \tag{3.1}$$

where $\chi(i\omega)$ is the same function as $\chi(s)$. This result is of course well known from electrical network theory.

For a direct proof we compute $b(s)$ for the above excitation. Thus, with

$$\begin{aligned} \hat{F}(s) &= \int_0^\infty dt \exp(-st)u(t)F_{0,\omega}\exp(i\omega t) \\ &= F_{0,\omega}/(s - i\omega) \end{aligned} \tag{3.2}$$

we have $b(s) = \chi(s)F_{0,\omega}/(s - i\omega)$, and for the response we obtain

$$\begin{aligned} \langle \Delta B(t)\rangle &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} ds \frac{\exp(st)\chi(s)F_{0,\omega}}{s - i\omega} \\ &= \frac{1}{2\pi i}\oint_C dw \frac{\exp(iwt)\chi(iw)F_{0,\omega}}{w - \omega}, \end{aligned} \tag{3.3}$$

where C is a contour which encircles all poles counterclockwise. *For a passive system $\chi(s)$ has no poles in the right half of the complex s-plane, or $\chi(iw)$ has no poles below the real axis of the complex frequency plane (causality principle).* Thus, letting $w_k = \omega_k + i\mu_k$ denote the poles, we have

$$\chi(iw) = \sum_k \frac{g_k}{w - (\omega_k + i\mu_k)}, \quad \mu_k \geq 0 \tag{3.4}$$

(for simplicity we assume single poles). From Cauchy's theorem we find

$$\langle \Delta B(t)\rangle = \exp(i\omega t)\chi(i\omega)F_{0,\omega} + \sum_k \frac{g_k \exp(i\omega_k t - \mu_k t)}{\omega_k - \omega + i\mu_k}. \tag{3.5}$$

The summand damps out for $\mu_k > 0$. (If $\mu_k = 0$, we consider an excitation $\lim_{\epsilon\to 0+} F_{0,\omega}\exp(i\omega t - \epsilon t)$. We thus arrive at (3.1).

The results of the s-plane of the previous section are now carried over in the frequency domain. The complex susceptance becomes the Fourier–Laplace transform (or one-sided Fourier transform) of the response function,

$$\chi_{BA}(i\omega) = \int_0^\infty dt \exp(-i\omega t)\phi_{BA}(t). \tag{3.6}$$

With ϕ_{BA} given by (2.13) this yields the commutator form

$$\chi_{BA}(i\omega) = (1/\hbar i)\int_0^\infty dt \exp(-i\omega t)\mathrm{Tr}\{[A, B(t)]\rho_{eq}\}. \tag{3.7}$$

Generally we write $\chi(i\omega) = \chi'(\omega) + i\chi''(\omega)$. From the causality principle one easily shows the Kramers–Kronig relations[28]

$$\chi'(\omega) = \frac{1}{\pi}P\int_{-\infty}^{\infty} \frac{\chi''(\omega')}{\omega' - \omega} d\omega', \tag{3.8a}$$

$$\chi''(\omega) = -\frac{1}{\pi}P\int_{-\infty}^{\infty} \frac{\chi'(\omega')}{\omega' - \omega} d\omega', \tag{3.8b}$$

indicating that χ' and χ'' are each other's Hilbert transform; here P denotes the principal value.

For the conductance we obtain likewise

$$L_{BA}(i\omega) = \int_0^\infty dt \exp(-i\omega t)\phi_{\dot{B}A}(t); \tag{3.9}$$

or

$$L_{BA}(i\omega) = (1/\hbar i)\int_0^\infty dt \exp(-i\omega t)\mathrm{Tr}\{[A, \dot{B}(t)]\rho_{eq}\}. \tag{3.10}$$

Again $L(i\omega) = L'(\omega) + iL''(\omega)$, with L' and L'' being each other's Hilbert transform.

The generalized susceptance can also be expressed in Ψ_{BA}. From (2.19) with integration by parts, one finds

TABLE I. Commutator forms.

Basic function	Commutator form		Derived quantity	
ϕ_{BA}	$\frac{1}{\hbar i}\mathrm{Tr}\{[A, B(t)]\rho_{eq}\}$	(2.13)	$\chi_{BA}(i\omega)=\int_0^\infty dt\exp(-i\omega t)\phi_{BA}(t)$	(3.6)
$\phi_{\dot{B}A}$	$\frac{1}{\hbar i}\mathrm{Tr}\{[A, \dot{B}(t)]\rho_{eq}\}$	(2.21)	$L_{BA}(i\omega)=\int_0^\infty dt\exp(-i\omega t)\phi_{\dot{B}A}(t)$	(3.9)
Ψ_{BA}	$\frac{1}{\hbar i}\int_t^\infty dt'\,\mathrm{Tr}\{[A, B(t')]\rho_{eq}\}$	(2.18)	$\chi_{BA}(i\omega)=\chi_{BA}^{dc}-i\omega\int_0^\infty dt\exp(-i\omega t)\Psi_{BA}(t)$	(3.11)
$\phi_{\dot{B}\dot{A}}$	$\frac{1}{\hbar i}\mathrm{Tr}\{[\dot{A}, \dot{B}(t)]\rho_{eq}\}$	(2.25)	$L_{BA}(i\omega)=\int_0^\infty dt\,\frac{\exp(-i\omega t)-1}{\hbar\omega}\phi_{\dot{B}\dot{A}}(t)$	(3.12)

$$\chi_{BA}(i\omega)=-\int_0^\infty dt\exp(-i\omega t)\frac{d}{dt}\Psi_{BA}(t)$$
$$=\Psi_{BA}(0)-i\omega\int_0^\infty dt\exp(-i\omega t)\Psi_{BA}(t)$$
$$=\chi_{BA}^{dc}-i\omega\int_0^\infty dt\exp(-i\omega t)\Psi_{BA}(t). \quad (3.11)$$

Finally, the complex conductance is expressible as a transform of $\phi_{\dot{B}\dot{A}}(t)$. From (2.30)

$$L_{BA}(i\omega)=-\int_0^\infty dt\,\frac{\exp(-i\omega t)-1}{i\omega}\phi_{\dot{B}\dot{A}}(t)$$
$$=\int_0^\infty dt\,\frac{\exp(-i\omega t)-1}{\hbar\omega}\mathrm{Tr}\{[\dot{A}, \dot{B}(t)]\rho_{eq}\}, \quad (3.12)$$

which is the standard expression (see, e.g., van Velsen[29]). We give a summary of the various formulas in Table I.

3.2. The Kubo form and the Fujita form

The commutator expressions of the previous subsection will be converted into correlation expressions. There is a difficulty, however, with the correlation expressions; these correlations should contain no steady state parts, but involve only the fluctuations [otherwise the Fourier transforms contain delta functions and the Fourier–Laplace transforms contain $\delta_+(t)$ functions]. The derivations go smoothly if these parts are separated off in the commutator form, prior to going to Kubo expressions. Thus, let

$$\Delta A(t)=A(t)-I\langle A\rangle, \quad \Delta B(t)=B(t)-I\langle B\rangle, \quad (3.13)$$

where I is the unit operator and $\langle A\rangle$ and $\langle B\rangle$ are the equilibrium average values, $\mathrm{Tr}\rho_{eq}A$ and $\mathrm{Tr}\rho_{eq}B$. For the commutator we have

$$[A, B(t)]=[\Delta A, \Delta B(t)], \quad (3.14)$$

as is easily verified. For the Heisenberg operator we have from (2.7),

$$\Delta B(t)=\exp(iHt/\hbar)B\exp(-iHt/\hbar)-I\langle B\rangle$$
$$=\exp(iHt/\hbar)B\exp(-iHt/\hbar)-\exp(iHt/\hbar)I\langle B\rangle$$
$$\times\exp(-iHt/\hbar)$$
$$=\exp(iHt/\hbar)\Delta B\exp(-iHt/\hbar). \quad (3.15)$$

Clearly, then, all formulas of Table I can be carried over *verbatim* for the delta operators. These changes are not trivial, since the trace of unbounded operators is not unambiguous; in fact, different results are obtained if this change in operators is not made here but is later incorporated in the Kubo expressions. We believe, however, that the previous results in terms of the delta operators give what is really desired, since Kubo's "natural motion" does not concern the equilibrium values, but the deviations from equilibrium [note in this respect also the definition of the response as $\langle\Delta B(t)\rangle$ in (2.4)]. Whereas we will carry the Δ's where required, they will be dropped where optional, which will turn out to be the case frequently.[30]

We now proceed to obtain the Kubo form. For $\phi_{BA}(t)$ of Eq. (2.13) we also write, using cyclic permutivity,

$$\phi_{BA}(t)=\frac{1}{\hbar i}\mathrm{Tr}\{[\Delta A, \Delta B(t)]\rho_{eq}\}$$
$$=-\frac{1}{\hbar i}\mathrm{Tr}\{[\Delta A, \rho_{eq}]\Delta B(t)\}. \quad (3.16)$$

We now use Kubo's identity

$$\frac{d}{d\beta'}[\exp(\beta' H)\Delta A\exp(-\beta' H)]$$
$$=\exp(\beta' H)[H, \Delta A]\exp(-\beta' H). \quad (3.17)$$

From this by integration from 0 to β and remultiplication by $\exp(-\beta H)$,

$$[\Delta A, \exp(-\beta H)]=-i\hbar\int_0^\beta d\beta'\exp(-\beta H)$$
$$\times\exp(\beta' H)[H, \Delta A]\exp(-\beta' H). \quad (3.18)$$

From the Heisenberg equation of motion [or from differentiation of (2.7)],

$$\frac{dA}{dt}=\frac{1}{i\hbar}[A(t), H], \quad (3.19)$$

or, since $dA/dt=d(\Delta A)/dt$, also

$$\frac{d\Delta A}{dt}=-\frac{1}{i\hbar}\exp(iHt/\hbar)[H, \Delta A]\exp(-iHt/\hbar). \quad (3.20)$$

Taking $t=-i\hbar\beta$, the rhs of (3.18) and (3.20) are similar; thus (3.18) yields the following lemma.

Lemma:

$$[\Delta A, \exp(-\beta H)]=-i\hbar\int_0^\beta d\beta'\exp(-\beta H)\,\Delta A(-i\hbar\beta'); \quad (3.21)$$

here $\Delta\dot{A}$ means $(d/dt)(\Delta A)$. Substitution of (3.21) into (3.16) gives the Kubo form

$$\phi_{BA}(t) = \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq} \Delta\dot{A}(-i\hbar\beta')\Delta B(t)\}. \tag{3.22}$$

Now we have $\Delta\dot{A} = \dot{A}$; thus for the trace we find

$$\begin{aligned}\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}(-i\hbar\beta')\Delta B(t)\} &= \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')B(t)\} \\ &\quad - \langle B\rangle \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta)\} \\ &= \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')B(t)\} \\ &\quad - \langle B\rangle\langle\dot{A}(-i\hbar\beta)\rangle. \end{aligned} \tag{3.23}$$

But from the Heisenberg equation of motion we find

$$\langle\dot{A}(t)\rangle = \mathrm{Tr}\{\rho_{eq}\dot{A}(t)\} = \frac{1}{\hbar i}\mathrm{Tr}\{\rho_{eq}[A(t), H]\} = 0, \tag{3.24}$$

since $[\rho_{eq}, H] = 0$.[31] Hence, in (3.22) the Δ's are optional and we also have

$$\phi_{BA}(t) = \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')\, B(t)\}. \tag{3.22'}$$

For χ_{BA} this gives

$$\chi_{BA}(i\omega) = \int_0^\infty \exp(-i\omega t)\, dt \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')B(t)\}. \tag{3.25}$$

From (3.22') by the change $B \to \dot{B}$ we also have

$$\phi_{\dot{B}A}(t) = \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')\dot{B}(t)\}. \tag{3.26}$$

Thus, for L_{BA} one finds the more symmetrical result

$$L_{BA}(i\omega) = \int_0^\infty \exp(-i\omega t)\, dt \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')\dot{B}(t)\}. \tag{3.27}$$

We now find the Kubo form for Ψ. From (2.18), (3.22), and stationarity,

$$\begin{aligned}\Psi_{BA}(t) &= \int_t^\infty dt' \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta\dot{A}(-t')\Delta B(i\hbar\beta')\} \\ &= [\int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta A(-t')\Delta B(i\hbar\beta')\}]_\infty^t. \end{aligned} \tag{3.28}$$

Assuming the validity of the mixing property (2.28), we have

$$\begin{aligned}&\lim_{t\to\infty} \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta A(-i\hbar\beta')\Delta B(t)\} \\ &\quad = \beta\langle\Delta A\rangle\langle\Delta B\rangle = 0. \end{aligned} \tag{3.29}$$

Thus we find

$$\Psi_{BA}(t) = \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta A(-i\hbar\beta')\Delta B(t)\}. \tag{3.30}$$

In this expression the Δ's must be kept.[32] The susceptance is found from Ψ according to (3.11),

$$\begin{aligned}\chi_{BA}(i\omega) &= \chi_{BA}^{dc} - i\omega \int_0^\infty dt \exp(-i\omega t) \\ &\quad \times \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta A(-i\hbar\beta')\Delta B(t)\}. \end{aligned} \tag{3.31}$$

We will also introduce the function $\Psi_{\dot{B}\dot{A}}$, defined by

$$\Psi_{\dot{B}\dot{A}}(t) \equiv \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta\dot{A}(-i\hbar\beta')\Delta\dot{B}(t)\}. \tag{3.32}$$

Though in thermal equilibrium $\Psi_{\dot{B}\dot{A}} = \phi_{\dot{B}A}$, as is found by comparison with (3.26), the introduction of $\Psi_{\dot{B}\dot{A}}$ is justified, since often we deal with a *quasiequilibrium* state, in which there are mean fluxes, i.e., $\langle\dot{A}(t)\rangle \neq 0$, while the fluctuations still relax by thermal random processes. The conductance is now given by

$$\begin{aligned}L_{BA}(i\omega) &= \int_0^\infty dt \exp(-i\omega t)\Psi_{\dot{B}\dot{A}}(t) \\ &= \int_0^\infty dt \exp(-i\omega t) \int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta\dot{A}(-i\hbar\beta')\Delta\dot{B}(t)\}. \end{aligned} \tag{3.33}$$

In summary, expressions (3.25) and (3.27) are the basic Kubo results, valid in thermal equilibrium. The relations (3.31) and (3.33) are also useful in a quasiequilibrium state. Moreover, these expressions serve to obtain the fluctuation–dissipation theorem, since it is the relaxation function Ψ which is most easily linked to the fluctuation spectrum. The various results are listed in Table II.

The Kubo form has the advantage that the *classical frequency limit* is easily found. By this we mean the following. The rhs in (3.30) depends on $|t + i\hbar\beta'|$, $0 \leq \beta' \leq \beta$, as follows from stationarity, (2.26). We consider frequencies for which $|i\hbar\beta| \ll t \sim 1/\omega$, or $\hbar\omega \ll 1/\beta$. These are called classical frequencies since the the quantum correction factor $\mathcal{E}(\omega, T)$ (next subsection) is unity in this range. Equation (3.30) then gives

$$\begin{aligned}\Psi_{BA}^{cl}(t) &= \beta \, \mathrm{Tr}\{\rho_{eq}\Delta A(0)\, \Delta B(t)\} \\ &= \beta\langle\Delta\dot{B}(t)\Delta\dot{A}(0)\rangle^{cl} = \beta\Phi_{BA}^{cl}, \end{aligned} \tag{3.34}$$

where Φ_{BA}^{cl} is the classical fluctuation–correlation function.

TABLE II. Kubo forms.

Basic function	Kubo form		Derived quantity	
ϕ_{BA}	$\int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')B(t)\}$	(3.22')	$\chi_{BA}(i\omega) = \int_0^\infty dt \exp(-i\omega t)\phi_{BA}(t)$	(3.25)
$\phi_{\dot{B}A}$	$\int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\dot{A}(-i\hbar\beta')\dot{B}(t)\}$	(3.26)	$L_{BA}(i\omega) = \int_0^\infty dt \exp(-i\omega t)\phi_{\dot{B}A}(t)$	(3.27)
Ψ_{BA}	$\int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta A(-i\hbar\beta')\Delta B(t)\}$	(3.30)	$\chi_{BA}(i\omega) = \chi_{BA}^{dc} - i\omega\int_0^\infty dt \exp(-i\omega t)\Psi_{BA}(t)$ [also quasiequil.,	(3.31)]
$\Psi_{\dot{B}\dot{A}}$	$\int_0^\beta d\beta' \, \mathrm{Tr}\{\rho_{eq}\Delta\dot{A}(-i\hbar\beta')\Delta\dot{B}(t)\}$	(3.32)	$L_{BA}(i\omega) = \int_0^\infty dt \exp(-i\omega t)\Psi_{\dot{B}\dot{A}}(t)$ [also quasiequil.,	(3.33)]

The Kubo form can be further transformed into the Fujita form.[33] Introducing the fictitious Hamiltonian $H''_A = H - \alpha\dot{A}$ for the scalar case, or $H''_A = H - \boldsymbol{\alpha}\cdot\dot{\mathbf{A}}$ for the vectorial case, one easily shows

$$\frac{\partial}{\partial\alpha_\nu}\frac{1}{H''-z} = \frac{1}{H''-z}\dot{A}_\nu\frac{1}{H''-z}. \tag{3.35}$$

By Cauchy's theorem,

$$\frac{\partial}{\partial\alpha_\nu}\exp(-\beta H'') = -\frac{1}{2\pi i}\oint dz\exp(-\beta z)\frac{\partial}{\partial\alpha_\nu}\frac{1}{H''-z}. \tag{3.36}$$

Substituting (3.35), we have by convolution

$$\begin{aligned}\frac{\partial}{\partial\alpha_\nu}&\exp(-\beta H'')\\ &= \int_0^\beta d\beta'\exp[-(\beta-\beta')H'']\dot{A}_\nu\exp(-\beta H'')\\ &= \int_0^\beta d\beta'\exp(-\beta H'')\dot{A}_\nu(-i\hbar\beta').\end{aligned} \tag{3.37}$$

Thus Kubo's expressions transform to

$$\chi_{BA}(i\omega) = \lim_{\alpha\to 0}\frac{\partial}{\partial\alpha}\int_0^\infty dt\exp(-i\omega t)\,\mathrm{Tr}\{\rho'' B(t)\}, \tag{3.38}$$

$$L_{B_\mu A_\nu}(i\omega) = \lim_{\alpha\to 0}\frac{\partial}{\partial\alpha_\nu}\int_0^\infty dt\exp(-i\omega t)\mathrm{Tr}\{\rho''\dot{B}_\mu(t)\}, \tag{3.39}$$

where

$$\rho'' = \exp(-\beta H'')/\mathrm{Tr}\exp(-\beta H''). \tag{3.40}$$

3.3. The correlation form

We now give the full correlation forms. The quantum mechanical fluctuation–correlation function is defined as the Hermitized product

$$\begin{aligned}\Phi_{BA}(t) &\equiv \langle\Delta B(t)\Delta A(0)\rangle = \tfrac{1}{2}\mathrm{Tr}\{\rho_{eq}[\Delta B(t),\Delta A]_+\}\\ &= \tfrac{1}{2}[\mathrm{Tr}\{\rho_{eq}\Delta B(t)\Delta A\} + \mathrm{Tr}\{\rho_{eq}\Delta A\Delta B(t)\}].\end{aligned} \tag{3.41}$$

Whereas the response function and the relaxation function are commutators, the correlation function is an anticommutator. These functions can be expressed into each other by the following relations:

$$\phi_{BA}(t) = \beta\int_{-\infty}^\infty dt'\,\Gamma(t-t')\Phi_{B\dot{A}}(t'), \tag{3.42}$$

$$\Psi_{BA}(t) = \beta\int_{-\infty}^\infty dt'\,\Gamma(t-t')\Phi_{BA}(t'), \tag{3.43}$$

where $\Gamma(t)$ is a kernel, even in t (see below). The possibility for the connections (3.42) and (3.43) is based on the following lemma.

Lemma: The two-sided Fourier transforms of the expectations $\langle\Delta C(t)\Delta D\rangle$ and $\langle\Delta D\Delta C(t)\rangle$ for any two operators C and D are related by

$$\begin{aligned}&\int_{-\infty}^\infty dt\exp(-i\omega t)\,\mathrm{Tr}\{\rho_{eq}\Delta C(t)\Delta D\}\\ &\quad= \exp(-\beta\hbar\omega)\int_{-\infty}^\infty dt\exp(-i\omega t)\,\mathrm{Tr}\{\rho_{eq}\Delta D\Delta C(t)\}.\end{aligned} \tag{3.44}$$

The proof goes as follows.[20] We consider the rhs:

$$\begin{aligned}&\int_{-\infty}^\infty dt\exp(-i\omega t)\,\mathrm{Tr}\{\rho_{eq}\Delta D\Delta C(t)\}\,\mathrm{Tr}[\exp(-\beta H)]\\ &\quad= \int_{-\infty}^\infty dt\exp(-i\omega t)\,\mathrm{Tr}\{\exp(-\beta H)\Delta D\exp(iHt/\hbar)\\ &\qquad\times\Delta C\exp(-iHt/\hbar)\}\\ &\quad= \int_{-\infty}^\infty dt\exp(-i\omega t)\,\mathrm{Tr}\{\exp(iHt/\hbar)\\ &\qquad\times\Delta C\exp(-iHt)/\hbar)\exp(-\beta H)\Delta D\}\\ &\quad= \exp(\beta\hbar\omega)\int_{-\infty}^\infty dt\exp[-i\omega(t-i\hbar\beta)]\\ &\qquad\times\mathrm{Tr}\{\exp(-\beta H)\exp[iH(t-i\hbar\beta)/\hbar]\\ &\qquad\times\Delta C\exp[-iH(t-i\hbar\beta)/\hbar]\Delta D\}\\ &\quad= \exp(\beta\hbar\omega)\int_{-\infty-i\hbar\beta}^{\infty-i\hbar\beta} dz\exp(-i\omega z)\,\mathrm{Tr}\{\exp(-\beta H)\Delta C(z)\,\Delta D\}.\end{aligned} \tag{3.45}$$

One now assumes analyticity on $0 \leq \mathrm{Im}z \leq \hbar\beta$ and one performs a contour integration along the real axis $-R \leq x \leq R$, the axis $z = x - i\hbar\beta$, and the axes $z = \pm R + iy$. Assuming the mixing property to be valid, the contribution along the latter axes vanishes since $\langle\Delta C\rangle = \langle\Delta D\rangle = 0$. Thus the result (3.44) follows.

From (2.13) and stationarity we have for the response function

$$d\phi_{BA}(t)/dt = -(1/\hbar i)\,\mathrm{Tr}\{\rho_{eq}[\Delta\dot{A},\Delta B(t)]\}. \tag{3.46}$$

Let $\hat{\phi}$ denote the two-sided Fourier transform; then

$$\begin{aligned}\hat{\phi}_{BA}(\omega) = \frac{1}{\hbar\omega}\int_{-\infty}^\infty dt\exp(-i\omega t)[&\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}\Delta B(t)\}\\ &- \mathrm{Tr}\{\rho_{eq}\Delta B(t)\Delta\dot{A}\}].\end{aligned} \tag{3.47}$$

Also let $\hat{\Phi}$ be the two-sided Fourier transform of the correlation function,

$$\begin{aligned}\hat{\Phi}_{B\dot{A}}(\omega) = \tfrac{1}{2}\int_{-\infty}^\infty dt\exp(-i\omega t)[&\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}\Delta B(t)\}\\ &+ \mathrm{Tr}\{\rho_{eq}\Delta B(t)\Delta\dot{A}\}].\end{aligned} \tag{3.48}$$

Then by the Lemma (3.44),

$$\hat{\phi}_{BA}(\omega) = [\mathcal{E}(\omega,T)]^{-1}\hat{\Phi}_{B\dot{A}}(\omega), \tag{3.49}$$

where

$$\mathcal{E}(\omega,T) = \frac{\hbar\omega}{2}\frac{1+\exp(-\beta\hbar\omega)}{1-\exp(-\beta\hbar\omega)} = \frac{\hbar\omega}{2}\coth\left(\frac{\beta\hbar\omega}{2}\right) \tag{3.50}$$

is the quantum noise correction factor (average energy of an harmonic oscillator at temperature T). By the convolution theorem of Fourier transforms, the inverse of (3.49) yields (3.42), with

$$\begin{aligned}\Gamma(t) &= \frac{1}{2\pi}\int_{-\infty}^\infty d\omega\exp(i\omega t)\frac{\tanh(\beta\hbar\omega/2)}{\beta\hbar\omega/2}\\ &= \frac{2}{\hbar\beta\pi}\log\coth\left(\frac{\pi}{2\beta\hbar}|t|\right).\end{aligned} \tag{3.51}$$

For the relaxation function the derivation goes similarly. Fourier transforming (2.19) one finds

$$\begin{aligned}\hat{\Psi}_{BA}(\omega) = \frac{1}{\hbar\omega}\int_{-\infty}^\infty dt\exp(-i\omega t)\\ \times[\mathrm{Tr}\{\rho_{eq}\Delta A\Delta B(t)\} - \mathrm{Tr}\{\rho_{eq}\Delta B(t)\Delta A\}];\end{aligned} \tag{3.52}$$

TABLE III. Correlation forms.

Correlation functions	Relation to ϕ or Ψ
$\Phi_{BA}^{\cdot}=\frac{1}{2}\mathrm{Tr}\{\rho_{eq}[\Delta B(t),\Delta\dot{A}]_+\}$	$\phi_{BA}(t)=\beta\int_{-\infty}^{\infty}dt'\,\Gamma(t-t')\Phi_{B\dot{A}}^{\cdot}(t')$
$\Phi_{BA}=\frac{1}{2}\mathrm{Tr}\{\rho_{eq}[\Delta B(t),\Delta A]_+\}$	$\Psi_{BA}(t)=\beta\int_{-\infty}^{\infty}dt'\,\Gamma(t-t')\Phi_{BA}(t')$
$\Phi_{\dot{B}\dot{A}}^{\cdot\cdot}=\frac{1}{2}\mathrm{Tr}\{\rho_{eq}[\Delta\dot{B}(t),\Delta\dot{A}]_+\}$	$\Psi_{\dot{B}\dot{A}}^{\cdot\cdot}(t)=\beta\int_{-\infty}^{\infty}dt'\,\Gamma(t-t')\Phi_{\dot{B}\dot{A}}^{\cdot\cdot}(t')$
$\chi_{BA}(i\omega)=\beta\int_0^{\infty}dt\exp(-i\omega t)\int_{-\infty}^{\infty}dt'\,\Gamma(t-t')\Phi_{B\dot{A}}^{\cdot}(t')$	
$\chi_{BA}(i\omega)-\chi_{BA}^{dc}=-i\omega\beta\int_0^{\infty}\exp(-i\omega t)\int_{-\infty}^{\infty}dt'\,\Gamma(t-t')\Phi_{BA}(t')$	
$L_{BA}(i\omega)=\beta\int_0^{\infty}dt\exp(-i\omega t)\int_{-\infty}^{\infty}dt'\,\Gamma(t-t')\Phi_{\dot{B}\dot{A}}^{\cdot\cdot}(t')$	

whence by the lemma (3.44),

$$\hat{\Psi}_{BA}(\omega)=[\mathcal{E}(\omega,T)]^{-1}\hat{\Phi}_{BA}(\omega). \tag{3.53}$$

The inverse Fourier transform yields immediately (3.43). We still note that for classical frequencies

$$\Gamma(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\exp(i\omega t)=\delta(t), \tag{3.54}$$

so that (3.43) goes over into (3.34).

We can now express the transport coefficients in $\Phi(t)$, by combining (3.6), (3.9), (3.11), or (3.33) with the relations (3.42) and (3.43). E.g., (3.6) leads to

$$\chi_{BA}(i\omega)=\beta\int_0^{\infty}dt\exp(-i\omega t)\int_{-\infty}^{\infty}dt'\Gamma(t-t')\Phi_{B\dot{A}}^{\cdot}(t'). \tag{3.55}$$

The various expressions are summarized in Table III.

The fluctuation–dissipation theorems are the Fourier transforms of the last relations in the table. For completeness, we briefly indicate the results. To obtain the transforms, it is most easy to go back to (3.53). By the Wiener–Khintchine theorem[34] the spectral density of the fluctuations $\langle\Delta\dot{B}\Delta\dot{A}\rangle$ is $2\hat{\Phi}_{BA}(\omega)$; thus from (3.53)

$$S_{BA}(\omega)=2\mathcal{E}(\omega,T)\int_{-\infty}^{\infty}dt\exp(-i\omega t)\Psi_{BA}(t). \tag{3.56}$$

Employing the transposition property (see Sec. 10), the integral on the rhs of (3.56) becomes

$$\begin{aligned}&\int_{-\infty}^{\infty}dt\exp(-i\omega t)\Psi_{BA}(t)\\&=\int_0^{\infty}dt\exp(-i\omega t)\Psi_{BA}(t)+\int_0^{\infty}dt\exp(i\omega t)\Psi_{AB}(t)\\&=2\int_0^{\infty}dt\cos\omega t[\Psi_{BA}(t)]^s-2i\int_0^{\infty}dt\sin\omega t[\Psi_{BA}(t)]^a,\end{aligned} \tag{3.57}$$

where we split into odd and even parts, forming the symmetric and antisymmetric expressions

$$[\Psi_{BA}]^s=\tfrac{1}{2}(\Psi_{BA}+\Psi_{AB}),\quad[\Psi_{BA}]^a=\tfrac{1}{2}(\Psi_{BA}-\Psi_{AB}). \tag{3.58}$$

We now use (3.11), *noticing its validity also for quasi-equilibrium*. Splitting this expression into real and imaginary parts and defining χ^a and χ^s similarly as in (3.58), noting $\chi_{AB}^{dc}=\chi_{BA}^{dc}$, we have

$$[\chi'_{BA}]^a=-\omega\int_0^{\infty}dt\sin\omega t[\Psi_{BA}(t)]^a, \tag{3.59a}$$

$$[\chi''_{BA}]^s=-\omega\int_0^{\infty}dt\cos\omega t[\Psi_{BA}(t)]^s. \tag{3.59b}$$

Upon substitution of (3.59) into (3.57), the theorem follows,

$$S_{BA}(\omega)=-4\mathcal{E}(\omega,T)\frac{1}{\omega}\{[\chi''_{BA}(\omega)]^s-i[\chi'_{BA}(\omega)]^a\}. \tag{3.60}$$

Likewise, we consider the spectrum $S_{\dot{B}\dot{A}}$, expressed in $\Psi_{\dot{B}\dot{A}}$ by a result similar to (3.56), and we split into even and odd parts. We further find the symmetric and antisymmetric tensors of L based on the quasiequilibrium expression (3.33). The result is the theorem

$$S_{\dot{B}\dot{A}}(\omega)=4\mathcal{E}(\omega,T)\{[L'_{BA}(\omega)]^s+i[L''_{BA}(\omega)]^a\}. \tag{3.61}$$

B. THE VAN HOVE LIMIT; REINTERPRETATION FOR THE CASE OF WEAK INTERACTION WITH A RESERVOIR

4. CRITICISM OF LINEAR RESPONSE THEORY

4.1. Van Kampen's criticism of the linearization

A critical discussion of the assumptions involved and of the validity of linear response theory has been given by van Kampen.[11,35] He states his uneasiness in the following way: "Linear response theory does provide expressions for the phenomenological coefficients, but I assert that *it arrives at these expressions by a mathematical exercise rather than by describing the actual mechanism which is responsible for the response.*" Indeed, we agree that the physical content of the theory, as it stands, is minimal.

Van Kampen's frontal assault concerns the linearization approximation. The two pertinent questions are: What does this linearization mean, and, how good is it?

First, let us return to Eqs. (2.2)–(2.6). Suppose Eq. (2.3) had not been iterated. It is easily found that instead of (2.6) we had then obtained the *exact result*

$$\begin{aligned}\langle\Delta B(t)\rangle&=(1/\hbar i)\,\mathrm{Tr}\int_0^t d\tau\,[A,B(t-\tau)]\rho(\tau)F(\tau)\\&=(1/\hbar i)\,\mathrm{Tr}\int_0^t d\tau[A,B(\tau)]\rho(t-\tau)F(t-\tau).\end{aligned} \tag{4.1}$$

The linearization amounts to replacing $\rho(t-\tau)\to\rho_{eq}$. The linearization therefore means that part of the memory of the system is erased. Thus, if the linearization is prolonged for large times, it has the effect of randomizing. As van Kampen states: "*The effect of randomization is simulated by the linear approximation.*"

The next question is, up to what field and for which time intervals the linearization is good, i.e., not subject to the above criticisms. Van Kampen makes the following estimate. A particle undergoes by the field a shift $d=\frac{1}{2}t^2eE/m^*$ (E is the electric field, e the electron charge, and m^* the effective mass in a solid lattice). This shift should be small compared to the diameter of the scatterers. Let the latter have a cross section of 100 Å and let $m^*=0.01m$, where m is the free electron mass; then the above gives $t^2E\lesssim 10^{-18}\,\mathrm{V\,s^2/cm}$. Taking for t an observation time of one second, van Kampen arrives at the condition $E\lesssim 10^{-18}$ V/cm!

In the present study we take a less pessimistic view. Clearly, the linearization approximation is reasonable, if we can greatly reduce the time interval t over which the linearization must work. For this reason we started the perturbation at $t=0$ rather than at $t=-\infty$. If we can further argue that a small relaxation time τ_r is *inherent* in the response function, then the above estimate is good if, say, $t=10\tau_r$. Letting $\tau_r \lesssim 10^{-11}$ s, we find that fields up to 100 V/cm are permissible. Thus, the acceptability of the linearization hinges critically on the emergence of (a) well-defined relaxation time(s).

4.2. Our criticism

In our opinion the main shortcomings of the general formalism as described in Sec. 2 and 3 are: (a) absence of dissipation; (b) insufficient conditions for relaxation; (c) ad hoc introduction of the canonical ensemble. We now discuss these points, together with a model that alleviates the objections.

(a) *Dissipation*: Linear response theory speaks of dissipation and associated transport coefficients, but *nowhere is the dynamics commensurate with dissipation introduced.* This is typical for any treatment based on the complete, microscopic motion of particles in the system. In a complete description, involving the coordinates of all particles and bodies, it is impossible to distinguish between "motion" and "scattering" since all forces are treated on the same mechanical basis. Dissipation, however, involves such a distinction, in that it requires that there be an agency which randomizes what is considered to be the motion proper.

It is also well known that in a microscopically complete description no entropy is produced. This is based on the fact that the derivative of the Gibbs entropy is zero, $\dot{S}_G = -k(d/dt)\mathrm{Tr}\rho\log\rho = 0$, as is found by substituting for $\dot{\rho}$ from von Neumann's equation. The problem of linear response theory is thus directly linked with the problem of explaining irreversibility on a mechanical basis. It is our opinion that, apart from such "tricks" as coarse graining, time smoothing, etc., *irreversibility stems from the physical stringency which calls for a partitioning of the Hamiltonian.* Thus we write,

$$H = H^0 + \lambda V, \tag{4.2}$$

where H^0 describes the part of interest ("motion proper") and λV the interaction causing dissipation. In the context of linear response theory λV is most readily interpreted as the interaction with a reservoir [see part (c) of this subsection], though the framework of the mathematics which follows (Secs. 6–8) remains the same if λV is an interaction acting within the system. In practice, we assume that H^0 can be diagonalized, either by exact separation of variables, or through some mean field (Hartree–Fock) procedure. The perturbation interaction λV then causes transitions among the states of H^0; or, in more classical terms, the interaction randomizes the motion contained in H^0, thus producing heat.[36]

Van Hove[37] gives several examples. For a lattice gas, H^0 contains the harmonic vibrations and λV the anharmonic forces (imperfect harmonic crystal); for a metal or semiconductor, H^0 contains the electron gas (if A and B are operators pertaining to the electrons only) and λV is the electron–phonon interaction. We note that H^0 is still a many-body Hamiltonian, which may contain several types of weak or strong interaction, such as electron– electron interaction in the second example cited.

Another way of looking at irreversibility is given by Fano.[25] He observes that the Liouville operator $\mathcal{L}$ in the complete description causes nothing but norm-conserved rotations of the density operator in Liouville space. However, as he states, "the variation of a projection of ρ within a subspace does not constitute a rotation." Thus, irreversibility is evident when the motion is viewed within a subspace of the Liouville space.

One may argue that in Kubo's theory there is the splitting off of the excitation Hamiltonian $-AF(t)$. However, dissipation is *not caused* by this excitation Hamiltonian, as seems to be implied in Callen and Welton's original derivation of the linear response result[5] (this derivation is therefore believed to be at variance with Kubo's formulation).

(b) *Relaxation*: We noted in subsection 4.1 that the theory in order to be valid must exhibit relaxation, i.e., the existence of characteristic time constants for the response and relaxation functions. First of all we observe that no relaxation occurs in a finite system. This is so because the Heisenberg operators are harmonic in t, cf. Eq. (2.11); thus $\phi_{BA}(t)$ has finite Fourier components for a set of discrete frequencies, indicating oscillatory behavior.

For the infinite system, the situation is more complex. We note that $B(t)$ is a unitary transformation of B, with t as a continuous parameter. According to Stone's theorem,[38] there is a spectral decomposition of (2.11), viz.,

$$\exp(i\mathcal{L}t) = \int_{-\infty}^{\infty} \exp(iut)\,dP(u) \sim \int_{-\infty}^{\infty} \exp(iut)p(u)\,du; \tag{4.3}$$

here $P(u)$ are the projectors and $p(u) = dP/du$. For an infinite system u becomes a continuous variable. There is of course no guarantee that the Fourier integral exists. If it exists, then the result damps out, for we have

$$\lim_{t\to\infty} \int_{-\infty}^{\infty} \exp(iut)\,p(u) = 0 \tag{4.4}$$

according to the Riemann–Lebesgue theorem.[39] However, the time scale involved is hard to estimate; generally, it is doubtful whether the limit (4.4) has anything to do with the *observable* approach towards equilibrium. As we will see in the present paper, the structure of the decomposition (4.3) will be radically altered for the partitioned Hamiltonian (Sec. 8.2).

(c) *The equilibrium density operator*: We noted that in the thermodynamic limit the microcanonical density operator could be replaced by a canonical density operator, resulting in the choice (2.5) for ρ_{eq}. If this is to be more than a mathematical artifice, we must physi-

cally change the system, so it becomes open and in interaction with a heat bath. The von Neumann equation and subsequent solution then still applies strictly to the closed composite system composed of the system plus bath. Thus, let the composite Hamiltonian be

$$H^c = H + H^b = H^0 + H^b + \lambda V, \tag{4.5}$$

where H^b is the bath Hamiltonian. The interaction Hamiltonian is as before a perturbation on H^0, but, in addition, couples to the bath coordinates. The solution for the response is again found to be (2.6) with ρ_{eq} replaced by the composite system density operator ρ^c_{eq} and with the Heisenberg operator

$$B(t) = \exp(iH^c t/\hbar)\, B \exp(-iH^c t/\hbar)$$
$$= \exp(iHt/\hbar) B \exp(-iHt/\hbar); \tag{4.6}$$

$B(t)$ is unaltered since B is an operator of the system which commutes with the bath Hamiltonian. For the computation of $\Delta B(t)$ we first take the trace over the bath states; we may make an adiabatic approximation for the bath coordinates still occurring in $B(t)$ through λV. Thus the only effect is on the density operator $\mathrm{Tr}_{(\mathrm{bath})}\rho^c_{eq} = \rho_{eq}$. It is well known that, if the composite system operator ρ^c_{eq} is that of a microcanonical ensemble, then ρ_{eq} for the system proper is the canonical density operator, providing the bath reservoir is very large. Elimination of the bath variables thus leads to the old results (2.6) and (2.7), with H given by (4.2) and with ρ_{eq} given by (2.5). The bath can further be discarded. The essential point here is that the use of the canonical ensemble density operator in Kubo's theory *already implied* that we relinquished a completely detailed microscopic description of all coordinates of the system, thus leading naturally to the partitioned Hamiltonian (4.2).

5. PRESENT PROCEDURE: LARGE SYSTEMS, WEAK INTERACTION

We will compute the Heisenberg operators in the large system and weak interaction limit. The calculation follows the procedure initiated by van Hove in 1955 in a paper which dealt with the derivation of the Pauli master equation.[16]

The large system limit entails the use of continuous quantum numbers for the characterization of the eigenstates of H^0, denoted by $|\gamma\rangle$. Often it is useful to choose the energy ϵ as part of this set of numbers, thus $|\gamma\rangle = |\epsilon\alpha\rangle$.

The weak interaction limit must be taken judiciously since as the interaction grows weaker, the time required to reach equilibrium becomes longer. Thus let $\tau_1, \tau_2, \cdots$, be the relaxation times which will emerge from the weak interaction limit of the Liouville operator $i\mathcal{L}$ and let τ_r be the largest relaxation time; the weak interaction or van Hove limit then means

$$\lambda \to 0, \quad \tau_r \to \infty, \quad \lambda^2 \tau_r \text{ finite}. \tag{5.1}$$

It should be understood that in practice it suffices to take λ small and τ_r "large." The first statement can be taken literally since λ is already a dimensionless scaling factor. The latter statement means that τ_r is large when scaled with respect to the time τ_t it takes to effectuate a transition. The time intervals t of interest for the process, on the other hand, are to be scaled with respect to τ_r, cf. Sec. 4.1, i.e., we consider intervals for which t/τ_r is bounded. Thus, (5.1) means more properly

$$\lambda \to 0, \quad t/\tau_t \to \infty, \quad \lambda^2 t \text{ finite}. \tag{5.2}$$

The time scaling implies that we can distinguish between two time ranges, viz.

$$\text{(i) } \tau_t \ll \Delta t \ll \tau_r, \quad \text{(ii) } t \gg \tau_t; \tag{5.3}$$

these will be referred to as the small time behavior and the large time behavior. It should be understood that the possibility of these time ranges is an important consequence of the condition of weak interaction. (For strong interaction, the relaxation times and the transition times can be of the same order of magnitude.) It is this feature which allows us to distinguish between hydrodynamic or macroscopic relaxation and microscopic motion. See also van Kampen[40] for a discussion of the underlying physical ideas.

Another problem concerns the amount of coarse graining involved in macroscopic operators. Van Kampen[40] has indicated that such operators for sufficiently large time intervals almost commute with the Hamiltonian. In particular, in the interaction picture we are dealing with macroscopic operators B which can be replaced by mutilated operators which are "smeared out" over the energy cells $\delta\epsilon$ of H^0, to the extent that $B \to B_d$, where B_d is diagonal in the representation of H^0. The van Hove limit will be carried out for the operators $B_d(t)$. This leads, however, to major difficulties in the Kubo formalism, which involves the commutators of the exact operators, like $[B(t), \rho_{eq}]$. The replacement $B \to B_d$ and the subsequent van Hove limit must therefore be incorporated at the appropriate stage in the theory; in this respect we will see that the Kubo form is much more useful than the commutator form. In particular there will be no problem if, in addition to the large system and weak interaction limits, we carry through the classical frequency limit defined by

$$\hbar\omega\beta \to 0, \text{ or } t/\hbar\beta \to \infty \ (\beta = 1/kT). \tag{5.4}$$

The word "limit" is again to be taken *cum grano salis*. In practice, it means that we take ω sufficiently small or t sufficiently large, when scaled with respect to $1/\hbar\beta$ or $\hbar\beta$, respectively.

6. THE REDUCTION OF THE HEISENBERG OPERATORS

6.1. The master operator

The eigenstates of H^0 are denoted as $|\gamma\rangle = |\epsilon\alpha\rangle$. Let $W_{\gamma\gamma''}$ be the transition probability $\gamma \to \gamma''$ caused by λV. Then, by the "golden rule,"

$$W_{\gamma\gamma''} = \frac{2\pi\lambda^2}{\hbar}\, \delta(\epsilon_\gamma - \epsilon_{\gamma''})\, |\langle\gamma|\, V\, |\gamma''\rangle|^2. \tag{6.1}$$

The strict conservation of energy expressed by (6.1) holds, of course, only in the limit $\lambda \to 0$. For finite λV the energy will be conserved to within this measure, so it is useful to introduce energy cells, $\epsilon^i \le \epsilon_\gamma \le \epsilon^i + \delta\epsilon$. Transitions occur between the states of an energy cell,

but different cells are unconnected. As to V, we assume that it has no diagonal matrix elements; if $\langle\gamma|V|\gamma\rangle \neq 0$, this part (i.e., the diagonal part V_d) will be incorporated in H^0.

Let $f(\gamma)$ be an arbitrary function labeled by γ. We then introduce the "master operator" M by

$$Mf(\gamma) \equiv -\sum_{\gamma''} [W_{\gamma,\gamma''} f(\gamma'') - W_{\gamma\gamma''} f(\gamma)]. \tag{6.2}$$

Note that *a priori* we do not assume knowledge of the master equation; at this point the above is simply a definition. With $f(\gamma)$ we now associate a diagonal operator $\hat{f}$ such that

$$\hat{f} = \sum_{\gamma} |\gamma\rangle f(\gamma)\langle\gamma| \quad \text{or} \quad \langle\gamma|\hat{f}|\gamma\rangle = f(\gamma). \tag{6.3}$$

The $\hat{f}$'s are elements of a subspace of the Liouville space referred to as the "diagonal Liouville space." We multiply (6.2) by $|\gamma\rangle\langle\gamma|$ and sum over all states. The left-hand side will by definition denote the master operator Λ_d acting in the diagonal Liouville space,

$$\begin{aligned}\Lambda_d \hat{f} &\equiv \sum_{\gamma} |\gamma\rangle\langle\gamma| Mf(\gamma) \\ &= -\sum_{\gamma\gamma''} [|\gamma\rangle\langle\gamma| W_{\gamma,\gamma''}\langle\gamma''|\hat{f}|\gamma''\rangle - |\gamma\rangle\langle\gamma| W_{\gamma\gamma''}\langle\gamma|\hat{f}|\gamma\rangle].\end{aligned} \tag{6.4}$$

Let

$$D_\gamma A \equiv \langle\gamma|A|\gamma\rangle \tag{6.5}$$

be the definition of the diagonal projection value; we then have for Λ_d,

$$\Lambda_d \equiv -\sum_{\gamma\gamma''} [|\gamma\rangle\langle\gamma| W_{\gamma,\gamma''} D_{\gamma''} - |\gamma\rangle\langle\gamma| W_{\gamma\gamma''} D_\gamma]. \tag{6.6}$$

We also define the unit operator of the diagonal Liouville space by $\mathbf{1}\hat{f} = \hat{f}$; clearly,

$$\mathbf{1} = \sum_{\gamma} |\gamma\rangle\langle\gamma| D_\gamma. \tag{6.7}$$

We now claim the following main result. In the van Hove limit (5.2), the reduced Heisenberg operators take the form

$$B_d^R(t) \equiv \lim_{\lambda} B_d^H(t) = \exp(-\Lambda_d t) B_d. \tag{6.8}$$

Here, B_d is the diagonal operator associated with B as described in Sec. 5 and $\lim_\lambda$ is a notation for the van Hove limit; as before the unbracketed operator B_d is the zero time or Schrödinger operator. We formally write $B = B_d + B_{nd}$, where B_{nd} is the nondiagonal part; we then have $\exp(-\Lambda_d t)B_{nd} = B_{nd}$ as is found by series expansion. Thus (6.8) also reads

$$B^R(t) \approx \exp(-\Lambda_d t)(B_d + B_{nd}) = \exp(-\Lambda_d t)B_d + B_{nd}. \tag{6.8'}$$

An exact result will be given in Part II.[18]

6.2. Proof for small Δt

We first prove (6.8) from straightforward perturbation theory for Δt small in the sense indicated by (5.3), case (i). Let $U(t)$ be the evolution operator. We must then calculate $B_d(\Delta t) = U^\dagger(\Delta t) B_d U(\Delta t)$. The evolution operator satisfies the integral equation (see, e.g., Messiah[41])

$$\frac{\hbar i \partial U}{\partial t} = (H^0 + \lambda V)U, \tag{6.9}$$

or in integral form

$$U(t, t_0) = 1 - (i/\hbar)\int_{t_0}^{t} (H^0 + \lambda V)U(\tau, t_0)d\tau. \tag{6.9'}$$

This can be solved to all orders of perturbation by iteration. Writing $U = \sum_i U^{(i)}$, one obtains the well-known results:

$$U^{(0)}(\Delta t) = \exp(-iH^0\Delta t/\hbar), \tag{6.10a}$$

$$U^{(1)}(\Delta t) = (\lambda/i\hbar)\int_0^{\Delta t} d\tau_1 \exp[-iH^0(\Delta t - \tau_1)/\hbar] \times V \exp(-iH^0\tau_1/\hbar), \tag{6.10b}$$

$$U^{(2)}(\Delta t) = (\lambda/i\hbar)^2 \int_0^{\Delta t} d\tau_2 \int_0^{\tau_2} d\tau_1 \exp[-iH^0(\Delta t - \tau_2)/\hbar] \times V \exp[-iH^0(\tau_2 - \tau_1)/\hbar] V \exp(-iH^0\tau_1/\hbar). \tag{6.10c}$$

Computing now $\langle\gamma|B(\Delta t)|\gamma'\rangle$ up to order λ^2, we have the following contributions.

(i) $$\langle\gamma|U^{(0)\dagger} B_d U^{(0)}|\gamma'\rangle = \exp[i(\epsilon_\gamma - \epsilon_{\gamma'})\Delta t/\hbar]\langle\gamma|B_d|\gamma'\rangle = \langle\gamma|B_d|\gamma\rangle\delta_{\gamma\gamma'}. \tag{6.11}$$

(ii) We consider

$$\begin{aligned}&\langle\gamma|U^{(1)\dagger} B_d U^{(1)}|\gamma'\rangle \\ &= \frac{\lambda^2}{\hbar^2}\int_0^{\Delta t} d\tau_1 \int_0^{\Delta t} d\tau_1' \exp(i\epsilon_\gamma \tau_1'/\hbar) \\ &\times \langle\gamma| V\{\exp[iH^0(\Delta t - \tau_1')/\hbar] B_d \\ &\times \exp[-iH^0(\Delta t - \tau_1)/\hbar]\} V|\gamma'\rangle \exp(-i\epsilon_{\gamma'}\tau_1/\hbar).\end{aligned} \tag{6.12}$$

The diagonal operator in $\{\ \}$ we call C. We now use van Hove's *functional rule*,[16] i.e., we write down a linear functional relationship in as yet unknown kernels X and Y,

$$\begin{aligned}\langle\gamma|VCV|\gamma'\rangle &= \langle\gamma|(VCV)_d|\gamma'\rangle + \langle\gamma|(VCV)_{nd}|\gamma'\rangle \\ &= \delta_{\gamma\gamma'}\sum_{\gamma''}\langle\gamma''|C|\gamma'\rangle X(\gamma'', \gamma) \\ &\quad + \sum_{\gamma''}\langle\gamma''|C|\gamma'\rangle Y(\gamma''; \gamma, \gamma').\end{aligned} \tag{6.13}$$

The reason for this splitting is that diagonal and nondiagonal terms behave basically very differently: The diagonal terms will survive, whereas the nondiagonal terms represent interference effects which die out.[42] We can also write (6.13) differently,

$$\begin{aligned}VCV &= (VCV)_d + (VCV)_{nd} \\ &= \sum_{\gamma} |\gamma\rangle\langle\gamma|\langle\gamma|(VCV)_d|\gamma\rangle + \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma'|\langle\gamma|(VCV)_{nd}|\gamma'\rangle \\ &= \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma|\langle\gamma'|C|\gamma'\rangle X(\gamma', \gamma) \\ &\quad + \sum_{\gamma\gamma'\gamma''} |\gamma\rangle\langle\gamma'|\langle\gamma''|C|\gamma''\rangle Y(\gamma''; \gamma, \gamma').\end{aligned} \tag{6.13'}$$

We use (6.13) to rewrite (6.12),

$$\begin{aligned}&\langle\gamma|U^{(1)\dagger} B_d U^{(1)}|\gamma'\rangle \\ &= \frac{\lambda^2}{\hbar^2}\delta_{\gamma\gamma'}\sum_{\gamma''}\int_0^{\Delta t} d\tau_1 \int_0^{\Delta t} d\tau_1' \\ &\times \exp[i(\epsilon_\gamma - \epsilon_{\gamma''})(\tau_1' - \tau_1)/\hbar]\langle\gamma''|B_d|\gamma''\rangle X(\gamma'', \gamma)\end{aligned}$$

$$+\frac{\lambda^2}{\hbar^2}\sum_{\gamma''}\int_0^{\Delta t}d\tau_1\int_0^{\Delta t}d\tau_1'\exp[i(\epsilon_\gamma-\epsilon_{\gamma''})\tau_1'/\hbar]$$
$$\times\exp[-i(\epsilon_{\gamma'}-\epsilon_{\gamma''})\tau_1/\hbar]\langle\gamma''|B_d|\gamma''\rangle Y(\gamma'';\gamma,\gamma'). \quad (6.14)$$

In evaluating the first line we use

$$\int_0^{\Delta t}d\tau_1\int_0^{\Delta t}d\tau_1'\exp[i(\epsilon_\gamma-\epsilon_{\gamma''})(\tau_1'-\tau_1)/\hbar]$$
$$=\left|\int_0^{\Delta t}\exp[i(\epsilon_\gamma-\epsilon_{\gamma''})\tau_1/\hbar]d\tau_1\right|^2$$
$$=\frac{\sin^2[(\epsilon_\gamma-\epsilon_{\gamma''})\Delta t/2\hbar]}{[(\epsilon_\gamma-\epsilon_{\gamma''})/2\hbar]^2}$$
$$\approx 2\pi\hbar\Delta t\,\delta(\epsilon_\gamma-\epsilon_{\gamma''}), \quad (6.15)$$

providing $(\epsilon_\gamma-\epsilon_{\gamma''})\Delta t/\hbar\gg 1$. We will thus identify the two limits on Δt, foreseen in (5.3)

$$\tau_t\approx\hbar/(\epsilon_\gamma-\epsilon_{\gamma''})=\hbar/\delta\epsilon \quad (6.16)$$

(which time also follows from Heisenberg's uncertainty principle $\delta t\,\delta\epsilon\approx\hbar$). For the upper limit on Δt we have the relaxation time

$$\tau_r\approx\hbar\delta\epsilon/\lambda^2V^2. \quad (6.17)$$

Thus Δt is "microscopically large" and "macroscopically small," when the range $\tau_t\ll\Delta t\ll\tau_r$ is satisfied. For the other part in (6.14) we note

$$\int_0^{\Delta t}d\tau_1'\exp[i(\epsilon_\gamma-\epsilon_{\gamma''})\tau_1'/\hbar]\sim 2\pi\delta_-[(\epsilon_\gamma-\epsilon_{\gamma''})/\hbar]$$
$$=\pi\hbar\delta(\epsilon_\gamma-\epsilon_{\gamma''})+\pi\hbar i\mathcal{P}/(\epsilon_\gamma-\epsilon_{\gamma''}), \quad (6.18)$$

and likewise

$$\int_0^{\Delta t}d\tau\exp[-i(\epsilon_{\gamma'}-\epsilon_{\gamma''})\tau_1/\hbar]$$
$$\sim 2\pi\delta_+[(\epsilon_{\gamma'}-\epsilon_{\gamma''})/\hbar]$$
$$=\pi\hbar\delta(\epsilon_{\gamma'}-\epsilon_{\gamma''})-\pi\hbar i\mathcal{P}/(\epsilon_{\gamma'}-\epsilon_{\gamma''}). \quad (6.19)$$

Since $\gamma\neq\gamma'$ in the second part, the delta functions multiply to zero, so that we do not have the anomaly of a square delta function. The important point is that the result of the second part of (6.14) goes with λ^2, while the first part goes with $\lambda^2\Delta t$. Thus in the van Hove limit the final result is

$$\lim_\lambda\langle\gamma|U^{(1)\dagger}B_dU^{(1)}|\gamma'\rangle$$
$$=\frac{2\pi\lambda^2\Delta t}{\hbar}\delta_{\gamma\gamma'}\sum_{\gamma''}\langle\gamma''|B_d|\gamma''\rangle\langle X(\gamma'',\gamma)\delta(\epsilon_{\gamma''}-\epsilon_\gamma). \quad (6.20)$$

(iii) Consider

$$\langle\gamma|U^{(0)\dagger}B_dU^{(1)}|\gamma'\rangle+\langle\gamma|U^{(1)\dagger}B_dU^{(0)}|\gamma'\rangle$$
$$=\langle\gamma|U^{(0)\dagger}B_dU^{(1)}|\gamma'\rangle+\langle\gamma'|U^{(0)\dagger}B_dU^{(1)}|\gamma\rangle^*. \quad (6.21)$$

One easily finds this goes with order λ; hence it vanishes in the van Hove limit.

(iv) Consider

$$S_{\gamma\gamma'}\equiv\langle\gamma|U^{(0)\dagger}B_dU^{(2)}|\gamma'\rangle+\langle\gamma'|U^{(0)\dagger}B_dU^{(2)}|\gamma\rangle^*$$
$$=\frac{-\lambda^2}{\hbar^2}\sum_{\gamma''\gamma'''}\int_0^{\Delta t}d\tau_2\int_0^{\tau_2}d\tau_1\exp(i\epsilon_\gamma\Delta t/\hbar)\langle\gamma|B_d|\gamma''\rangle$$
$$\times\exp[-i\epsilon_{\gamma''\gamma'''}(\Delta t-\tau_2)/\hbar]\langle\gamma''|V\exp[-iH^0(\tau_2-\tau_1)/\hbar]$$
$$\times V|\gamma'\rangle\exp(-i\epsilon_{\gamma'}\tau_1/\hbar)+\text{hcj}, \quad (6.22)$$

where hcj denotes the Hermititian conjugate, $i\to -i$, $\gamma\to\gamma'$. Now again use (6.13). This yields (notice there is no Y part for this case):

$$S_{\gamma\gamma'}=\frac{-\lambda^2}{\hbar^2}\sum_{\gamma''\gamma'''}\int_0^{\Delta t}d\tau_2\int_0^{\tau_2}d\tau_1\langle\gamma|B_d|\gamma'\rangle X(\gamma'',\gamma')$$
$$\times\exp[-i(\epsilon_{\gamma'}-\epsilon_\gamma)\Delta t/\hbar]\exp[i(\tau_2-\tau_1)(\epsilon_{\gamma'}-\epsilon_{\gamma''})/\hbar]+\text{hcj}$$
$$=\frac{-2\lambda^2}{\hbar^2}\delta_{\gamma\gamma'}\,\mathrm{Re}\sum_{\gamma''\gamma'''}\langle\gamma|B_d|\gamma\rangle X(\gamma'',\gamma)$$
$$\times\int_0^{\Delta t}d\tau_2\int_0^{\tau_2}d\tau_1\exp[i(\tau_2-\tau_1)(\epsilon_\gamma-\epsilon_{\gamma''})/\hbar]. \quad (6.23)$$

The time integral is found to be

$$\text{integral}=\frac{\sin^2[(\epsilon_\gamma-\epsilon_{\gamma''})\Delta t/2\hbar]}{[(\epsilon_\gamma-\epsilon_{\gamma''})/2\hbar]^2}=2\pi\hbar\Delta t\,\delta(\epsilon_\gamma-\epsilon_{\gamma''}). \quad (6.24)$$

Hence,

$$\lim_\lambda S_{\gamma\gamma'}=\frac{-2\pi\lambda^2\Delta t}{\hbar}\delta_{\gamma\gamma'}\sum_{\gamma''}\langle\gamma|B_d|\gamma\rangle X(\gamma'',\gamma)\delta(\epsilon_\gamma-\epsilon_{\gamma''}). \quad (6.25)$$

Finally collecting terms, from (6.11), (6.20), and (6.25),

$$\lim_\lambda\langle\gamma|B_d(\Delta t)|\gamma'\rangle$$
$$=\langle\gamma|B_d|\gamma'\rangle+\frac{2\pi\lambda^2\Delta t}{\hbar}\delta_{\gamma\gamma'}\sum_{\gamma''}\langle\gamma''|B_d|\gamma''\rangle X(\gamma'',\gamma)$$
$$\times\delta(\epsilon_\gamma-\epsilon_{\gamma''})-\frac{2\pi\lambda^2\Delta t}{\hbar}\delta_{\gamma\gamma'}$$
$$\times\sum_{\gamma''}\langle\gamma|B_d|\gamma\rangle X(\gamma'',\gamma)\delta(\epsilon_\gamma-\epsilon_{\gamma''}). \quad (6.26)$$

Since the above holds for any diagonal B_d, we apply (6.26) in particular for $B_d=|\gamma^0\rangle\langle\gamma^0|$ (and we take $\gamma=\gamma'$). The lhs of (6.26) is then $|\langle\gamma|U^\dagger(\Delta t)|\gamma^0\rangle|^2$. Consider now the general solution for the diagonal part of ρ in the homogeneous von Neumann equation $\partial\rho/\partial t+(1/\hbar i)[\rho,H]=0$,

$$p(\gamma^0,t)\equiv\langle\gamma^0|\rho(t)|\gamma^0\rangle$$
$$=\sum_{\gamma,\gamma'}\langle\gamma^0|U(t)|\gamma'\rangle\langle\gamma'|\rho(0)|\gamma\rangle\langle\gamma|U^\dagger(t)|\gamma^0\rangle. \quad (6.27)$$

We make an *initial random phase assumption*, i.e., $\rho(0)$ is diagonal $\langle\gamma'|\rho(0)|\gamma\rangle=p(\gamma,0)\delta_{\gamma\gamma'}$. Then (6.27) reduces to

$$p(\gamma^0,t)=\sum_\gamma|\langle\gamma|U^\dagger(t)|\gamma^0\rangle|^2p(\gamma,0), \quad (6.28)$$

indicating that the conditional probability is to be identified as

$$|\langle\gamma|U^\dagger(t)|\gamma^0\rangle|^2=P(\gamma^0,t|\gamma,0). \quad (6.29)$$

Hence, from (6.26) for $B_d=|\gamma^0\rangle\langle\gamma^0|$,

$$P(\gamma^0,\Delta t|\gamma,0)=\frac{2\pi\lambda^2\Delta t}{\hbar}X(\gamma^0,\gamma)\delta(\epsilon_\gamma-\epsilon_{\gamma^0})$$
$$+\delta_{\gamma\gamma^0}\left(1-\frac{2\pi\lambda^2\Delta t}{\hbar}\sum_{\gamma''}X(\gamma'',\gamma)\,\delta(\epsilon_\gamma-\epsilon_{\gamma''}).\right. \quad (6.30)$$

For ease in notation, we interchange γ and γ^0. Thus (6.30) is consistent with the Stosszahlansatz

$$P(\gamma, \Delta t|\gamma^0, 0) = W_{\gamma\gamma}\Delta t + \delta_{\gamma\gamma^0}\left[1 - \sum_{\gamma''} W_{\gamma''\gamma}\Delta t\right], \quad (6.31)$$

where

$$W_{\gamma\gamma'} = \frac{2\pi\lambda^2}{\hbar} X(\gamma, \gamma')\delta(\epsilon_\gamma - \epsilon_{\gamma'}) = W_{\gamma'\gamma} \quad (6.32)$$

(property of microscopic reversibility). The identification of X follows from comparison of (6.32) with the "golden rule" (6.1),

$$X(\gamma, \gamma') = |\langle\gamma|V|\gamma'\rangle|^2. \quad (6.33)$$

Now we rewrite (6.26) for general B_d using (6.32). The Kronecker delta is rewritten, so that for any $F(\gamma)$,

$$\begin{aligned}\delta_{\gamma\gamma'}F(\gamma) &= \sum_{\bar\gamma} \delta_{\gamma\bar\gamma}\delta_{\bar\gamma\gamma'}F(\bar\gamma)\\ &= \sum_{\bar\gamma}\langle\gamma|\bar\gamma\rangle\langle\bar\gamma|\gamma'\rangle F(\bar\gamma)\\ &= \langle\gamma|\{\sum_{\bar\gamma}|\bar\gamma\rangle\langle\bar\gamma|F(\bar\gamma)\}|\gamma'\rangle. \end{aligned} \quad (6.34)$$

We then easily obtain

$$\begin{aligned}\lim_\lambda \langle\gamma|B_d(\Delta t)|\gamma'\rangle &= \langle\gamma|B_d|\gamma'\rangle + \langle\gamma|\{\sum_{\bar\gamma\gamma''}|\bar\gamma\rangle\langle\bar\gamma|W_{\gamma''\bar\gamma}\Delta t D_{\gamma''}B_d\}|\gamma'\rangle\\ &\quad - \langle\gamma|\{\sum_{\bar\gamma\gamma''}|\bar\gamma\rangle\langle\bar\gamma|W_{\bar\gamma\gamma''}\Delta t D_{\bar\gamma}B_d\}|\gamma'\rangle. \end{aligned} \quad (6.35)$$

We note that the rhs contains the master operator Λ_d of Eq. (6.6). Thus the final result is

$$\lim_\lambda \langle\gamma|B_d(\Delta t)|\gamma'\rangle = \langle\gamma|B_d|\gamma'\rangle - \langle\gamma|\Lambda_d\Delta t B_d|\gamma'\rangle, \quad (6.36)$$

or also

$$B_d^R(\Delta t) = B_d - \Lambda_d\Delta t B_d, \quad (6.37)$$

which is the small time form of (6.8).

6.3. Large time properties

A simple "derivation" is as follows. Equation (6.36) gives

$$[B_d^R(\Delta t) - B_d]/\Delta t = -\Lambda_d B_d. \quad (6.38)$$

Here $B_d = B_d^R(0)$. In the limit $\Delta t \to 0$ we have $dB_d/dt = -\Lambda_d B_d$. Thus by integration

$$B_d^R(t) = \exp(-\Lambda_d t)B. \quad (6.39)$$

This "derivation" is no good, however, since at the beginning of each time interval Δt we must reenact the random phase assumption (*repeated random phase assumption*). This is incompatible with quantum mechanical evolution, as is easily seen from von Neumann's equation,

$$\frac{\partial\rho}{\partial t} = \frac{i}{\hbar}[\rho, H^0] + \frac{i\lambda}{\hbar}[\rho, V]; \quad (6.40)$$

if ρ is diagonal at all times, then $[\rho, H^0] = 0$, thus ρ is then determined by λV alone, which is clearly wrong. We therefore return to van Hove's paper, for a more profound evaluation of $\langle\gamma|B_d(t)|\gamma'\rangle$ for all times t.

Prior to proceeding to the proof, we shall first examine the meaning of the exponential operator. Consider the equation

$$\frac{\partial f(\gamma, t)}{\partial t} + \sum_{\gamma''}[-W_{\gamma''\gamma}f(\gamma') + W_{\gamma\gamma''}f(\gamma)] = Q(\gamma, t). \quad (6.41)$$

The formal solution is (see, e.g., Morse and Feshbach,[43] Sec. 7.3)[44]

$$\begin{aligned} f(\gamma, t) &= \int_0^{t+0} dt' \sum_{\gamma'} g(\gamma, t; \gamma', t')Q(\gamma', t')\\ &= \int_0^{t+0} dt' \int_{\text{all states}} \Delta\gamma'\, g(\gamma, t; \gamma', t')Q(\gamma', t'). \end{aligned} \quad (6.42)$$

Here $\Delta\gamma' = Z(\gamma')\,d\gamma'$, with $Z(\gamma')$ being the density of states. The Green's function satisfies

$$\begin{aligned}\frac{\partial g(\gamma, t; \gamma', t')}{\partial t} + \sum_{\gamma''}[-W_{\gamma''\gamma}g(\gamma'', t; \gamma', t') + W_{\gamma\gamma''}g(\gamma, t; \gamma', t')]\\ = \delta(t - t')\delta_{\gamma\gamma'} = \delta(t - t')\delta(\gamma - \gamma')/Z(\gamma). \end{aligned} \quad (6.43)$$

Again, we associate with any function $f(\gamma, t)$ an operator $\hat f(t)$, see (6.3) (the dependence on the parameter t is of no importance). We then define the Green's operator, being a superoperator in Liouville space, by

$$\mathcal{G}(t, t')\hat f(t') = \sum_\gamma\sum_{\gamma'}|\gamma\rangle\, g(\gamma, t; \gamma', t')\langle\gamma|f(\gamma', t'). \quad (6.44)$$

Thus we also have

$$\mathcal{G}(t, t') = \sum_{\gamma\gamma'}|\gamma\rangle\, g(\gamma, t; \gamma', t')\langle\gamma|D_{\gamma'}. \quad (6.44')$$

Note also that the matrix element of $\mathcal{G}(t, t')\hat f$ behaves as a normal Green's operator in function space,

$$\begin{aligned}\langle\bar\gamma|\mathcal{G}(t, t')\hat f(t')|\bar{\bar\gamma}\rangle &= \delta_{\bar\gamma\bar{\bar\gamma}}\int\Delta\gamma' g(\bar\gamma, t; \gamma', t')f(\gamma', t')\\ &= \delta_{\bar\gamma\bar{\bar\gamma}}\int\Delta\gamma' g(\bar\gamma, t; \gamma', t')\langle\gamma'|\hat f(t')|\gamma'\rangle. \end{aligned} \quad (6.45)$$

Since the time axis can be shifted, we also define the single time Green's function

$$g(\gamma, t; \gamma') \equiv g(\gamma, t; \gamma', 0) = g(\gamma, t + \tau; \gamma', \tau) \quad (6.46)$$

and similarly the single time Green's operator

$$\mathcal{G}(t) \equiv \mathcal{G}(t, 0) = \mathcal{G}(t + \tau, \tau). \quad (6.47)$$

We now multiply (6.43) by $|\gamma\rangle\langle\gamma|f(\gamma', t')$ and we sum over γ and γ'. Then, comparing with (6.4) and (6.44), we find that $g(t, t')$ satisfies the simple operator equation

$$\frac{d\mathcal{G}(t, t')}{dt} + \Lambda_d\mathcal{G}(t, t') = \mathbf{I}\delta(t - t'), \quad (6.48)$$

where $\mathbf{I}$ is the unit operator (6.7).

The solution is

$$\mathcal{G}(t, t') = u(t - t')\exp[-\Lambda_d(t - t')] \quad (6.49)$$

or also

$$\mathcal{G}(t) = \exp(-\Lambda_d t) \quad (t > 0). \quad (6.50)$$

The Eq. (6.8) which we seek to prove therefore also reads

$$B_d^R(t) = \mathcal{G}(t)B_d, \quad (6.51)$$

or for the matrix elements, using (6.45)

$$\langle\gamma|B_d^R(t)|\gamma'\rangle=\delta_{\gamma\gamma'}\int\Delta\gamma''\,g(\gamma,t;\gamma'')\langle\gamma''|B_d|\gamma''\rangle. \quad (6.52)$$

This is therefore the relationship we seek to establish.

7. REDUCTION FOR LARGE TIMES

7.1. Derivation and proof via the master equation

The calculation of $|\langle\gamma|U(t)|\gamma'\rangle|^2$ was carried out by van Hove in his 1955 paper. The computation of the matrix elements $\langle\gamma|B(t)|\gamma'\rangle$ can be carried out with the same summation technique. This was actually already noted by van Hove, though the computation as such was not given. We have therefore reconstructed this computation, using the splitting of U in diagonal and nondiagonal parts, as indicated in van Hove's paper. The details are given in Appendices A and B. First, $U_d(t)$ is found; then $U_{nd}(t)$ is expressed in U_d as a convergent series $\sum_{n=1}^{\infty}U_{nd}^{(n)}(t)$. Terms of order of magnitude $\lambda^2 t$ are carried along to all powers. The cross terms of U_d and U_{nd} obviously yield no diagonal terms; cross terms from $U_{nd}^{\dagger(n)}$ and $U_{nd}^{(m)}$ with $n\neq m$ are found to vanish in the weak coupling limit. One thus obtains, with $\bar\lambda=\lambda\sqrt{\hbar}$ and with Γ defined by

$$\Gamma(\gamma)=\pi\int\Delta\beta\,X(\beta,\gamma)\delta(\epsilon_\beta-\epsilon_\gamma), \quad (7.1)$$

the following matrix element for $B_d^R(t)$:

$$\langle\gamma|B_d^R(t)|\gamma'\rangle=\langle\gamma|U_d^\dagger(t)B_dU_d(t)|\gamma'\rangle+\sum_{n=1}^{\infty}\langle\gamma|U_{nd}^{(n)\dagger}(t)B_dU_{nd}^{(n)}(t)|\gamma'\rangle, \quad (7.2)$$

or

$$\begin{aligned}\langle\gamma|B_d^R(t)|\gamma'\rangle&=\delta_{\gamma\gamma'}u(t)\exp[-2\bar\lambda^2 t\Gamma(\gamma)]\langle\gamma|B_d|\gamma\rangle\\&+\delta_{\gamma\gamma'}\sum_{n=1}^{\infty}(\pi\bar\lambda^2)^n\int_0^{2t}d\theta_n\int_0^{\theta_n}d\theta_{n-1}\cdots\\&\times\int_0^{\theta_2}d\theta_1\int\Delta\beta_n\int\Delta\beta_{n-1}\cdots\int\Delta\beta_1\\&\times\delta(\epsilon_{\beta_n}-\epsilon_{\beta_{n-1}})\cdots\delta(\epsilon_{\beta_1}-\epsilon_\gamma)\\&\times X(\beta_n,\beta_{n-1})\cdots X(\beta_1,\gamma)\exp[-\lambda^{-2}\Gamma(\beta_n)(2t-\theta_n)]\\&\times\exp[-\bar\lambda^2\Gamma(\beta_{n-1})(\theta_n-\theta_{n-1})]\cdots\\&\times\exp[-\bar\lambda^2\Gamma(\beta_1)(\theta_2-\theta_1)]\exp[-\bar\lambda^2\Gamma(\gamma)\theta_1]\\&\times\langle\beta_n|B_d|\beta_n\rangle. \quad (7.3)\end{aligned}$$

[Note: in the term with $n=1$, factors with β_0 (not defined) are to be omitted.] In particular, take $B(t)=|\bar\gamma\rangle\langle\bar\gamma|$ and set $\gamma=\gamma'$; (7.3) then yields $P(\bar\gamma,t|\gamma')$. One finds

$$\begin{aligned}P(\bar\gamma,t|\gamma')&=\delta_{\bar\gamma\gamma'}u(t)\exp[-2\bar\lambda^2 t\,\Gamma(\bar\gamma)]+\sum_{n=2}^{\infty}(\pi\bar\lambda^2)^n\\&\times\int_0^{2t}d\theta_n\int_0^{\theta_n}d\theta_{n-1}\cdots\int_0^{\theta_2}d\theta_1\int\Delta\beta_{n-1}\cdots\int\Delta\beta_1\\&\times\delta(\epsilon_{\bar\gamma}-\epsilon_{\beta_{n-1}})\delta(\epsilon_{\beta_{n-1}}-\epsilon_{\beta_{n-2}})\cdots\\&\times\delta(\epsilon_{\beta_1}-\epsilon_{\gamma'})X(\bar\gamma,\beta_{n-1})X(\beta_{n-1},\beta_{n-2})\cdots X(\beta_1,\gamma')\\&\times\exp[-\bar\lambda^2\Gamma(\bar\gamma)(2t-\theta_n)]\exp[-\bar\lambda^2\Gamma(\beta_{n-1})(\theta_n-\theta_{n-1})]\cdots\\&\times\exp[-\bar\lambda^2\Gamma(\beta_1)(\theta_2-\theta_1)]\exp[-\bar\lambda^2\Gamma(\gamma')\theta_1]\\&+\pi\bar\lambda^2\int_0^{2t}d\theta_1\,\delta(\epsilon_{\bar\gamma}-\epsilon_{\gamma'})X(\bar\gamma,\gamma')\\&\times\exp[-\bar\lambda^2\Gamma(\bar\gamma)(2t-\theta_1)]\exp[-\bar\lambda^2\Gamma(\gamma')\theta_1] \quad (7.4)\end{aligned}$$

(the last term was written separately since the summand has too many undefined symbols, were it carried to $n=1$; for $n=2$, factors involving β_0 are to be omitted). Substituting (7.4) back into (7.3), we find the expected result

$$\langle\gamma|B_d^R(t)|\gamma'\rangle=\delta_{\gamma\gamma'}\int\Delta\bar\gamma\,P(\bar\gamma,t|\gamma)\langle\bar\gamma|B_d|\bar\gamma\rangle. \quad (7.5)$$

We now differentiate (7.4),

$$\begin{aligned}\frac{\partial P(\bar\gamma,t|\gamma')}{\partial t}&=-2\bar\lambda^2\Gamma(\bar\gamma)P(\bar\gamma,t|\gamma')+\delta_{\bar\gamma\gamma'}\delta(t)\\&+2\sum_{n=2}^{\infty}(\pi\bar\lambda^2)^n\int_0^{2t}d\theta_{n-1}\cdots\int_0^{\theta_2}d\theta_1\int\Delta\beta_{n-1}\cdots\int\Delta\beta_1\\&\times\delta(\epsilon_{\bar\gamma}-\epsilon_{\beta_{n-1}})\delta(\epsilon_{\beta_{n-1}}-\epsilon_{\beta_{n-2}})\cdots\\&\times\delta(\epsilon_{\beta_1}-\epsilon_{\gamma'})X(\bar\gamma,\beta_{n-1})X(\beta_{n-1},\beta_{n-2})\cdots X(\beta_1,\gamma')\\&\times\exp[-\bar\lambda^2\Gamma(\beta_{n-1})(2t-\theta_{n-1})]\cdots\\&\times\exp[-\bar\lambda^2\Gamma(\beta_1)(\theta_2-\theta_1)]\exp[-\bar\lambda^2\Gamma(\gamma')\theta_1]\\&+2\pi\bar\lambda^2\delta(\epsilon_{\bar\gamma}-\epsilon_{\gamma'})X(\bar\gamma,\gamma')\exp[-2t\bar\lambda^2\Gamma(\gamma')]. \quad (7.6)\end{aligned}$$

Substituting (7.1) into the first term of (7.6) we notice that

$$\text{first term of (7.6)}=-\int\Delta\gamma''\,W_{\gamma'',\bar\gamma}P(\bar\gamma,t|\gamma'). \quad (7.7)$$

In the sum terms of (7.6) we change the summation index

$$\sum_{n=2}^{\infty}f_n=\sum_{n=2}^{\infty}f_{n+1}+f_2; \quad (7.8)$$

for the new integration variable β_n we write γ''. Hence,

sum term of (7.6) + last term

$$\begin{aligned}&=2\pi\bar\lambda^2\sum_{n=2}^{\infty}(\pi\bar\lambda^2)^n\int_0^{2t}d\theta_n\int_0^{\theta_n}d\theta_{n-1}\cdots\\&\times\int_0^{\theta_2}d\theta_1\int\Delta\gamma''\int\Delta\beta_{n-1}\cdots\int\Delta\beta_1\\&\times\delta(\epsilon_{\bar\gamma}-\epsilon_{\gamma''})\delta(\epsilon_{\gamma''}-\epsilon_{\beta_{n-1}})\delta(\epsilon_{\beta_{n-1}}-\epsilon_{\beta_{n-2}})\cdots\delta(\epsilon_{\beta_1}-\epsilon_{\gamma'})\\&\times X(\bar\gamma,\gamma'')X(\gamma'',\beta_{n-1})X(\beta_{n-1},\beta_{n-2})\cdots X(\beta_1,\gamma')\\&\times\exp[-\bar\lambda^2\Gamma(\gamma'')(2t-\theta_n)]\exp[-\bar\lambda^2\Gamma(\beta_{n-1})(\theta_n-\theta_{n-1})]\cdots\\&\times\exp[-\bar\lambda^2\Gamma(\beta_1)(\theta_2-\theta_1)]\exp[-\bar\lambda^2\Gamma(\gamma')\theta_1]\\&+2\pi\bar\lambda^2\cdot\pi\bar\lambda^2\int_0^{2t}d\theta_1\int\Delta\beta_1\delta(\epsilon_{\bar\gamma}-\epsilon_{\beta_1})\\&\times\delta(\epsilon_{\beta_1}-\epsilon_{\gamma'})X(\bar\gamma,\beta_1)X(\beta_1,\gamma')\\&\times\exp[-\bar\lambda^2\Gamma(\beta_1)(2t-\theta_1)]\exp[-\bar\lambda^2\Gamma(\gamma')\theta_1]\\&+2\pi\bar\lambda^2\delta(\epsilon_{\bar\gamma}-\epsilon_{\gamma'})X(\bar\gamma,\gamma')\exp[-2t\bar\lambda^2\Gamma(\gamma')]. \quad (7.9)\end{aligned}$$

Comparison with (7.4) shows that

$$(7.9) = 2\pi\bar{\lambda}^2 \int \Delta\gamma'' X(\bar{\gamma}, \gamma'')\delta(\epsilon_{\bar{\gamma}} - \epsilon_{\gamma''})P(\gamma'', t|\gamma')$$

$$= \int \Delta\gamma'' W_{\bar{\gamma}\gamma''} P(\gamma'', t|\gamma'). \quad (7.10)$$

Thus, from (7.6)–(7.10) we find, still using microscopic reversibility,

$$\frac{\partial P(\gamma, t|\gamma')}{\partial t} - \int \Delta\gamma'' [W_{\gamma\gamma''} P(\gamma'', t|\gamma') - W_{\gamma''\gamma} P(\gamma, t|\gamma')]$$

$$= \delta(t)\delta_{\gamma\gamma'} \quad (7.11)$$

or

$$\frac{\partial P(\gamma, t|\gamma')}{\partial t} + MP(\gamma, t|\gamma') = \delta(t)\delta(\gamma - \gamma')/Z(\gamma), \quad (7.11')$$

which is the *master equation*. Comparison with (6.43) shows

$$P(\gamma, t|\gamma') \equiv g(\gamma, t; \gamma'). \quad (7.12)$$

We need a few more properties of (6.43). Let $-\partial/\partial t + \tilde{M}$ be the adjoint operator. We note that $\tilde{M} = M$, for the integral operator is self-adjoint. Let $\tilde{g}(\gamma, t; \gamma', t')$ be the adjoint Green's function, i.e., the Green's function of $-\partial/\partial t + \tilde{M}$. We can associate with $\tilde{g}$ the conditional probability for the reverse time behavior, i.e., of $P(\gamma, -t|\gamma') = |\langle\gamma|U^\dagger|\gamma'\rangle|^2$, where $U^\dagger = U(-t)$. Thus,[45]

$$P(\gamma, -t|\gamma') = \tilde{g}(\gamma, -\gamma; \gamma'). \quad (7.13)$$

Since we also have[45]

$$P(\gamma, t|\gamma') = |\langle\gamma|U(t)|\gamma'\rangle|^2$$

$$= |\langle\gamma'|U^\dagger(t)|\gamma\rangle|^2 = P(\gamma', -t|\gamma), \quad (7.14)$$

comparison with (7.12) and (7.13) shows

$$g(\gamma, t; \gamma') = \tilde{g}(\gamma', -t; \gamma). \quad (7.15)$$

This is just the condition for reciprocity, since by shifting the time axis (7.15) also reads

$$g(\gamma, t; \gamma', 0) = \tilde{g}(\gamma', 0; \gamma, t). \quad (7.15')$$

Finally, since $\tilde{M} = M$, the adjoint Green's function for negative time satisfies the same equation as the regular Green's function for positive time,

$$\tilde{g}(\gamma', -t; \gamma) \equiv g(\gamma', t; \gamma). \quad (7.16)$$

Thus with (7.15)

$$g(\gamma, t; \gamma') = g(\gamma', t; \gamma), \quad (7.17)$$

indicating symmetry for fixed t. When (7.17) and (7.12) are substituted into (7.5) we arrive at

$$\langle\gamma|B_d^R(t)|\gamma'\rangle = \delta_{\gamma\gamma'} \int \Delta\bar{\gamma}\, g(\gamma, t; \bar{\gamma})\langle\bar{\gamma}|B_d|\bar{\gamma}\rangle, \quad (7.18)$$

which is (6.52). Q.E.D.

7.2. Proof via a differential equation for the matrix elements

Equation (7.17) indicates there is also symmetry in P,

$$P(\gamma, t|\gamma') = P(\gamma', t|\gamma). \quad (7.19)$$

This property can be shown to hold for the result (7.4). In the last term we simply change the integration variable, $\theta_1 \to 2t - \theta_1$. From the big summand we lift out the time integrals; with $2t = T$, let

$$\mathcal{J}_c \equiv \int_0^T d\theta_n \int_0^{\theta_n} d\theta_{n-1} \cdots \int_0^{\theta_2} d\theta_1$$

$$\times \exp[-\bar{\lambda}^2\Gamma(\bar{\gamma})(T - \theta_n)] \exp[-\bar{\lambda}^2\Gamma(\beta_{n-1})(\theta_n - \theta_{n-1})] \cdots$$

$$\times \exp[-\bar{\lambda}^2\Gamma(\beta_1)(\theta_2 - \theta_1)] \exp[-\bar{\lambda}^2\Gamma(\gamma')\theta_1]. \quad (7.20)$$

This integral is of the same form as the configurational partition function in Takahashi's nearest neighbor gas;[46] the functions Γ correspond with the two-body potentials! Thus, as there, we introduce the new variables

$$\begin{aligned} \tau_n &= T - \theta_n, \\ \tau_{n-1} &= \theta_n - \theta_{n-1}, \\ &\cdot \\ &\cdot \\ &\cdot \\ \tau_1 &= \theta_2 - \theta_1. \end{aligned} \quad (7.21)$$

The integral becomes

$$\mathcal{J}_c = \int_0^T d\tau_n \exp[-\bar{\lambda}^2\Gamma(\bar{\gamma})\tau_n] \int_0^{T-\tau_n} d\tau_{n-1}$$

$$\times \exp[-\bar{\lambda}^2\Gamma(\beta_{n-1})\tau_{n-1}] \cdots \int_0^{T-\tau_n-\cdots-\tau_2} d\tau_1$$

$$\times \exp[-\bar{\lambda}^2\Gamma(\beta_1)] \exp[-\bar{\lambda}^2\Gamma(\gamma')(T - \textstyle\sum_i \tau_i)], \quad (7.22)$$

which is a convolution of $n+1$ quantities. Thus we introduce the Laplace transforms

$$Z(s) = \int_0^\infty \exp(-sT)\mathcal{J}_c(T)\, dT,$$

$$F[s, \Gamma(\alpha)] = \int_0^\infty \exp[-sT - \bar{\lambda}^2\Gamma(\alpha)T]\, dT$$

$$= [s + \bar{\lambda}^2\Gamma(\alpha)]^{-1}. \quad (7.23)$$

The transformed result (7.22) becomes

$$Z(s) = F[s, \Gamma(\bar{\gamma})]F[s, \Gamma(\gamma')] \prod_{k=1}^{n-1} F[s, \Gamma(\beta_k)]. \quad (7.24)$$

We note the symmetry in $\bar{\gamma}$ and γ'. We thus interchange $\bar{\gamma}$ and γ' and we change the β's cyclically, $\beta_1 \to \beta_2, \cdots, \beta_{n-1} \to \beta_1$, and we transform (7.24) back with the factors F in the prescribed order. The new $\mathcal{J}_c$ is put into (7.4) and the integration variables $\beta_1 \cdots \beta_{n-1}$ of this expression are changed cyclically. The symmetry then follows.

With this symmetry knowledge, we will now show that the matrix elements (7.3) can be handled directly, without appeal to the master equation. Evaluating the time integrals with Laplace transforms in a similar way as above, we find that we can interchange γ and β_n everywhere, except in $\Delta\beta_n$ and in the matrix element $\langle\beta_n|B_d|\beta_n\rangle$; this yields the following alternate expression,

$$\langle\gamma|B_d^R(t)|\gamma'\rangle$$

$$= \delta_{\gamma\gamma'} u(t) \exp[-2\bar{\lambda}^2 t\Gamma(\gamma)]\langle\gamma|B_d|\gamma\rangle$$

$$+ \delta_{\gamma\gamma'} \sum_{n=2}^{\infty} (\pi\bar{\lambda}^2)^n \int_0^{2t} d\theta_n \int_0^{\theta_n} d\theta_{n-1} \cdots$$

$$\times \int_0^{\theta_2} d\theta_1 \int \Delta\beta_n \int \Delta\beta_{n-1} \cdots \int \Delta\beta_1$$

$$\times \delta(\epsilon_\gamma - \epsilon_{\beta_{n-1}})\delta(\epsilon_{\beta_{n-1}} - \epsilon_{\beta_{n-2}}) \cdots$$

$$\times\delta(\epsilon_{\beta_1}-\epsilon_{\beta_n})X(\gamma,\beta_{n-1})X(\beta_{n-1},\beta_{n-2})\cdots X(\beta_1,\beta_n)$$
$$\times\exp[-\bar{\lambda}^2\Gamma(\gamma)(2t-\theta_n)]\exp[-\bar{\lambda}^2\Gamma(\beta_{n-1})(\theta_n-\theta_{n-1})]\cdots$$
$$\times\exp[-\bar{\lambda}^2\Gamma(\beta_1)(\theta_2-\theta_1)]\exp[-\bar{\lambda}^2\Gamma(\beta_n)\theta_1]\langle\beta_n|B_d|\beta_n\rangle$$
$$+2\pi\bar{\lambda}^2\int_0^{2t}d\theta_1\int\Delta\beta_1\delta(\epsilon_{\beta_1}-\epsilon_\gamma)X(\beta_1,\gamma)$$
$$\times\exp[-\bar{\lambda}^2\Gamma(\gamma)(2t-\theta_1)]\exp[-\bar{\lambda}^2\Gamma(\beta_1)\theta_1]\langle\beta_1|B_d|\beta_1\rangle. \tag{7.25}$$

If this is differentiated, one easily arrives at the differential equation, using the same steps as in the previous subsection,

$$\frac{\partial\langle\gamma|B_d^R(t)|\gamma\rangle}{\partial t}+\int\Delta\gamma''[-W_{\gamma\cdot\cdot,\gamma}\langle\gamma''|B_d^R(t)|\gamma''\rangle+W_{\gamma\gamma\cdot\cdot}\langle\gamma|B_d^R(t)|\gamma\rangle$$
$$=\delta(t)\langle\gamma|B_d|\gamma\rangle, \tag{7.26}$$

which is of the form (6.41). The solution is by (6.42),

$$\langle\gamma|B_d^R(t)|\gamma\rangle=\int\Delta\gamma''\,g(\gamma,t;\gamma'')\langle\gamma''|B_d|\gamma''\rangle. \tag{7.27}$$

Since the off-diagonal elements are zero—see (7.25)—we once more obtain the result (6.52). Q.E.D.

8. APPROACH TO EQUILIBRIUM

In this section we first treat some known results concerning the master equation and the nonequilibrium distribution function (subsection 8.1). Next we establish new results for the evolution of the reduced Heisenberg operators $B_d^R(t)$ (subsection 8.2). The two problems are of course related; the first aspect looks essentially at the equilibrium approach in the Schrödinger frame, while the other aspect views the situation within the (reduced) Heisenberg frame.

8.1. Properties of the operator M and relaxation of the distribution function

In the space of real functions $f(\gamma)$ the scalar product is defined with the density of states $Z(\gamma)$, as a weighting factor,

$$(f,g)\equiv\int d\gamma\,Z(\gamma)f(\gamma)g(\gamma)=\int\Delta\gamma f(\gamma)g(\gamma), \tag{8.1}$$

with the integral over all γ-space. The master operator M of (6.2) is self-adjoint,

$$(g,Mf)=-\int\int\Delta\gamma\,\Delta\gamma''\{W_{\gamma\cdot\cdot,\gamma}g(\gamma)f(\gamma'')-W_{\gamma\gamma''}g(\gamma)f(\gamma)\}$$
$$=-\int\int\Delta\gamma\Delta\gamma''\{W_{\gamma\cdot\cdot,\gamma}g(\gamma'')f(\gamma)-W_{\gamma\gamma\cdot\cdot}g(\gamma)f(\gamma)\}$$
$$=(f,Mg), \tag{8.2}$$

where in the first integral we interchanged γ and γ'' and used microscopic reversibility, $W_{\gamma\gamma\cdot\cdot}=W_{\gamma\cdot\cdot,\gamma}$. The master operator M is positive semidefinite. We have for any f,

$$(f,Mf)=\int\int\Delta\gamma\Delta\gamma''W_{\gamma\gamma\cdot\cdot}f(\gamma)[f(\gamma)-f(\gamma'')]$$
$$=\int\int\Delta\gamma\Delta\gamma''W_{\gamma\gamma\cdot\cdot}f(\gamma'')[f(\gamma'')-f(\gamma)]$$
$$=\tfrac{1}{2}\int\int\Delta\gamma\Delta\gamma''W_{\gamma\gamma\cdot\cdot}[f(\gamma)-f(\gamma'')]^2\geq 0; \tag{8.3}$$

in the last line we took half the sum of the preceding expressions.

We must now be a bit more specific on the structure of M. As noticeable from (6.1), M contains a delta function in the energy. We substitute (6.32) and integrate out the delta function,

$$M^\epsilon f(\epsilon,\alpha)=-\frac{2\pi\lambda^2}{\hbar}\int\Delta\alpha''X(\epsilon,\alpha\alpha'')[f(\epsilon,\alpha'')-f(\epsilon,\alpha)]. \tag{8.4}$$

Clearly, $M=M^\epsilon$ *only acts on* α *space,* ϵ *being a parameter.* We define the scalar product in this subspace,

$$(f,g)^\epsilon\equiv\int\Delta\alpha f(\epsilon,\alpha)g(\epsilon,\alpha), \tag{8.5}$$

with $\Delta\alpha=Z(\epsilon,\alpha)\,d\alpha$ being the number of states of energy ϵ, having α within $(\alpha,\alpha+d\alpha)$. The connection with the full scalar product in γ space is

$$(f,g)=\int_0^\infty d\epsilon\,(f,g)^\epsilon. \tag{8.6}$$

Property (8.3) reads in particular

$$(f,M^\epsilon f)^\epsilon=\frac{\pi\lambda^2}{\hbar}\iint\Delta\alpha\Delta\alpha''X(\epsilon,\alpha\alpha'')[f(\epsilon,\alpha)-f(\epsilon,\alpha'')]^2$$
$$\geq 0, \tag{8.7}$$

since $X>0$ [see (6.33)].

Next, we consider the eigenfunctions and eigenvalues defined by

$$M^\epsilon\varphi_\kappa(\epsilon,\alpha)=\kappa(\epsilon)\varphi_\kappa(\epsilon,\alpha),\quad(\varphi_\kappa,\varphi_{\kappa'})^\epsilon=\delta_{\kappa\kappa'}. \tag{8.8}$$

The eigenvalues are positive semidefinite as is found by substituting $f=\varphi_\kappa$ in (8.7),

$$\kappa=\frac{\pi\lambda^2}{\hbar}\iint\Delta\alpha\Delta\alpha''X(\epsilon,\alpha\alpha'')[\varphi_\kappa(\epsilon,\alpha)-\varphi_\kappa(\epsilon,\alpha'')]^2$$
$$\geq 0. \tag{8.9}$$

There is an eigenvalue $\kappa=0$, for the equation $M^\epsilon\varphi_0=0$ has a solution, viz., the equilibrium solution of the master equation. The corresponding eigenfunction is independent of α for fixed energy; using the normalization we find

$$\varphi_0(\epsilon,\alpha)=\varphi_0(\epsilon,\alpha'')=1/\sqrt{G(\epsilon)}, \tag{8.10}$$

where

$$G(\epsilon)=\int Z(\epsilon,\alpha)\,d\alpha \tag{8.11}$$

is the degeneracy of the energy ϵ or, if coarse graining is taken into account, the number of states in an energy cell $(\epsilon,\epsilon+\delta\epsilon)$. It can further be shown that the eigenvalue $\kappa=0$ of a discrete spectrum is nondegenerate since otherwise M is reducible, i.e., the master equation splits into two or more master equations. For the same reason, the eigenvalue $\kappa=0$ must be isolated if part of the spectrum is continuous. Thus the spectrum is always of the form

$$\kappa=0\ \text{(equilibrium distirubution)}$$
$$\text{other }\kappa\geq\delta>0. \tag{8.12}$$

From (8.9) and (6.33) the estimate $\delta \sim \lambda^2 V^2/\hbar\delta\epsilon$, where $\delta\epsilon$ is a measure for the width of the energy distribution, indicated in (6.17) is evident.

We now consider $P(\gamma, t|\gamma')$ as given by the master equation (7.11'). Integrating out the energy delta function in M, we find that P is separable as $P(\gamma, t|\gamma') = P^\epsilon(\epsilon\alpha, t|\epsilon\alpha')\delta(\epsilon - \epsilon')$ with for P^ϵ the more restricted master equation,

$$\frac{\partial P^\epsilon(\epsilon\alpha, t|\epsilon\alpha')}{\partial t} + M^\epsilon P^\epsilon(\epsilon\alpha, t|\epsilon\alpha') = \delta(t)\delta(\alpha - \alpha')/Z(\epsilon, \alpha). \tag{8.13}$$

Expanding P^ϵ in eigenfunctions $\varphi_\kappa(\epsilon, \alpha)$ one easily finds

$$P^\epsilon(\epsilon\alpha, t|\epsilon\alpha') = \sum_\kappa \exp(-\kappa t)\varphi_\kappa(\epsilon\alpha)\varphi_\kappa(\epsilon\alpha'), \quad \kappa = \kappa(\epsilon), \tag{8.14}$$

where S denotes a sum for the discrete spectrum, and may contain integral parts for the continuous spectrum. For the nonconditional distribution we have

$$\begin{aligned} p(\epsilon\alpha, t) &= \int \Delta\gamma' P(\gamma, t|\gamma')p(\gamma', 0) \\ &= \int d\epsilon' \int \Delta\alpha' P^\epsilon(\epsilon\alpha, t|\epsilon\alpha')\delta(\epsilon - \epsilon')p(\epsilon'\alpha', 0) \\ &= \sum_\kappa \exp(-\kappa t)\int \Delta\alpha'\varphi_\kappa(\epsilon\alpha)\varphi_\kappa(\epsilon\alpha')p(\epsilon\alpha', 0). \end{aligned} \tag{8.15}$$

For $t \to \infty$ this yields in view of (8.12) and (8.10)

$$p_{eq}(\epsilon\alpha) = \frac{1}{G(\epsilon)}\int \Delta\alpha' p(\epsilon\alpha', 0) = \frac{\rho(\epsilon, 0)}{G(\epsilon)}, \tag{8.16}$$

where ρ is the energy distribution function. Clearly due to mixing of α states within an energy cell, $p_{eq}(\epsilon\alpha)$ becomes independent of α.[47] The microcanonical distribution is direct. The canonical distribution follows only if the distribution over the various energy cells at $t = 0$, $\rho(\epsilon, 0)$, was already canonical. The nonmixing of states from different energy cells is of course due to the (artificial) assumption of vanishing λV in the van Hove limit.

Finally, the computation of the Gibb's entropy requires us to find

$$S_G = -k \int\int d\epsilon\, \Delta\alpha\, p(\epsilon\alpha, t)\log p(\epsilon\alpha, t). \tag{8.17}$$

For $p(\epsilon\alpha, t)$ one finds by integrating (8.13), over all initial distributions,

$$\frac{\partial p(\epsilon\alpha, t)}{\partial t} + M^\epsilon p(\epsilon\alpha, t) = \delta(t)p(\epsilon\alpha, 0). \tag{8.18}$$

[This equation is sometimes also called the master equation; it has the same structure as (7.26).] From this one easily obtains

$$\frac{dS_G}{dt} = \frac{\pi k\lambda^2}{\hbar}\int d\epsilon \iint \Delta\alpha\Delta\alpha'' X(\epsilon, \alpha\alpha'')[p(\epsilon\alpha'', t) - p(\epsilon\alpha, t)] \times [\log p(\epsilon\alpha'', t) - \log p(\epsilon\alpha, t)] \geq 0. \tag{8.19}$$

The dissipative nature of the system is thus now abundantly demonstrated.

8.2. PROPERTIES OF THE OPERATOR Λ AND RELAXATION OF THE REDUCED HEISENBERG OPERATORS

In the Liouville space the scalar product of two operators is generally defined as[48]:

$$\{C, D\} = \mathrm{Tr}\, CD^\dagger = \int\int \Delta\gamma\Delta\gamma'\langle\gamma|C|\gamma'\rangle\langle\gamma|D|\gamma'\rangle^*. \tag{8.20}$$

We observe that $\{C, D\}^* = \{D, C\} = \{C^\dagger, D^\dagger\}$. In the subspace of Hermitian diagonal operators (diagonal Liouville space) the product (8.20) gives for operators $\hat{f}$ as defined in (6.3),

$$\{\hat{f}, \hat{g}\} = \int \Delta\gamma\langle\gamma|\hat{f}|\gamma\rangle\langle\gamma|\hat{g}|\gamma\rangle = (f, g). \tag{8.21}$$

We will also be considering more restricted operators denoted by an overbar,

$$\bar{f}(\epsilon) = \int \Delta\alpha\, |\epsilon\alpha\rangle f(\epsilon, \alpha)\langle\epsilon\alpha|, \tag{8.22a}$$

$$f(\epsilon, \alpha)\delta(\epsilon - \epsilon') = \langle\epsilon'\alpha|\bar{f}(\epsilon)|\epsilon'\alpha\rangle. \tag{8.22b}$$

We consider the operator Λ_d acting in the diagonal Liouville space. From its definition (6.6) one finds that Λ_d decomposes into a sum of operators for each energy cell,

$$\Lambda_d = -\frac{2\pi\lambda^2}{\hbar}\int_0^\infty d\epsilon \iint \Delta\alpha\Delta\alpha''\, |\epsilon\alpha\rangle\langle\epsilon\alpha|X(\epsilon, \alpha\alpha'') \times (D^\epsilon_{\alpha''} - D^\epsilon_\alpha) \approx \sum_i \int \Delta\alpha\, |\epsilon^i\alpha\rangle\langle\epsilon^i\alpha|/\!\!/^{\epsilon^i} = \sum_i \Lambda^i_d, \tag{8.23}$$

where we consider the energy divided over energy cells ϵ^i when convenient; $/\!\!/$ is defined such that

$$/\!\!/^\epsilon \bar{f}(\epsilon) = M^\epsilon f(\epsilon, \alpha). \tag{8.24}$$

Theorem 1: The eigenvalues and eigenvectors of Λ^i_d are $\kappa^i \equiv \kappa(\epsilon^i)$ and $\bar{\varphi}_{\kappa i}$, respectively, where

$$\bar{\varphi}_{\kappa i} = \int \Delta\alpha\, |\epsilon^i\alpha\rangle\varphi_{\kappa i}(\epsilon^i, \alpha)\langle\epsilon^i\alpha|, \tag{8.25}$$

in which κ^i and $\varphi_{\kappa i}(\epsilon^i, \alpha)$ are the eigenvalues and the eigenfunctions of M^{ϵ^i}, respectively, see (8.8).

The proof is direct:

$$\begin{aligned} \Lambda^i_d\bar{\varphi}_{\kappa i} &= \int \Delta\alpha\, |\epsilon^i\alpha\rangle\langle\epsilon^i\alpha|/\!\!/^{\epsilon^i}\bar{\varphi}_{\kappa i} \\ &= \int \Delta\alpha\, |\epsilon^i\alpha\rangle\langle\epsilon^i\alpha|M^{\epsilon^i}\varphi_{\kappa i}(\epsilon^i, \alpha) \\ &= \kappa^i \int \Delta\alpha\, |\epsilon^i\alpha\rangle\langle\epsilon^i\alpha|\varphi_{\kappa i}(\epsilon^i, \alpha) = \kappa^i\bar{\varphi}_{\kappa i}. \end{aligned} \tag{8.26}$$

Theorem 2: The eigenvalues and eigenvectors of Λ_d are all $\{\kappa^i\}$ and all $\{\bar{\varphi}_{\kappa i}\}$, respectively, $0 \leq \epsilon^i \leq \infty$.

The proof is straightforward. From (8.22b) we find that $/\!\!/^{\epsilon^j}\bar{f}(\epsilon^i) = 0$, $i \neq j$. It follows that $\Lambda^j_d\bar{\varphi}_{\kappa i} = 0$. Thus

$$\Lambda_d\bar{\varphi}_{\kappa i} = \sum_k \Lambda^k_d\bar{\varphi}_{\kappa i} = \Lambda^i_d\bar{\varphi}_{\kappa i} = \kappa^i\bar{\varphi}_{\kappa i}. \tag{8.27}$$

Theorem 3: We have the following projector property,

$$P_{\kappa i}\hat{f} = (f(\epsilon^i, \alpha), \varphi_{\kappa i}(\epsilon^i, \alpha))^{\epsilon^i}\bar{\varphi}_{\kappa i}, \tag{8.28}$$

where $P_{\kappa i}$ is the projector associated with $\bar{\varphi}_{\kappa i}$.

The proof is most easily given in Dirac notation:

$$P_{\kappa^i}\hat{f} = |\bar{\varphi}_{\kappa^i}\rangle\{\bar{\varphi}_{\kappa^i}|\hat{f}\} = |\bar{\varphi}_{\kappa^i}\} \operatorname{Tr}(\hat{f}\bar{\varphi}_{\kappa^i}); \tag{8.29}$$

now

$$\begin{aligned}\operatorname{Tr}(\hat{f}\bar{\varphi}_{\kappa^i}) &= \sum_{\epsilon} \int \Delta\alpha \langle \epsilon\alpha|\hat{f}|\epsilon\alpha\rangle\langle\epsilon\alpha|\bar{\varphi}_{\kappa^i}|\epsilon\alpha\rangle \\ &= \sum_{\epsilon} \int \Delta\alpha f(\epsilon,\alpha)\varphi_{\kappa^i}(\epsilon^i,\alpha)\delta_{\epsilon\epsilon^i} \\ &= (f(\epsilon^i,\alpha), \varphi_{\kappa^i}(\epsilon^i,\alpha))^{\epsilon^i},\end{aligned} \tag{8.30}$$

which completes the proof.

We thus have the following decomposition theorem,

$$\Lambda_d = \sum_i \mathop{S}_{\kappa^i} \kappa^i P_{\kappa^i}; \tag{8.31}$$

going over to a continuous energy scale we also have

$$\exp(-\Lambda_d t) = \int_0^\infty d\epsilon \mathop{S}_{\kappa(\epsilon)} \exp(-\kappa(\epsilon)t) P_{\kappa(\epsilon)}. \tag{8.32}$$

Finally, using (8.28) we have for the reduced Heisenberg operator

$$\begin{aligned}B_d^R(t) &= \exp(-\Lambda_d t)B \\ &= \int_0^\infty d\epsilon \mathop{S}_{\kappa(\epsilon)} \exp[-\kappa(\epsilon)t] \int \Delta\alpha \langle\epsilon\alpha|B|\epsilon\alpha\rangle \\ &\quad \times \varphi_{\kappa(\epsilon)}(\epsilon,\alpha)\bar{\varphi}_{\kappa(\epsilon)}.\end{aligned} \tag{8.33}$$

The expectation value as function of t is now found to be, employing (8.22b),

$$\begin{aligned}\langle B_d^R(t)\rangle &= \operatorname{Tr}\{\rho_{eq} B_d^R(t)\} \\ &= \int_0^\infty d\epsilon \mathop{S}_{\kappa(\epsilon)} \exp[-\kappa(\epsilon)t] \\ &\quad \times \int \Delta\alpha \langle\epsilon\alpha|B|\epsilon\alpha\rangle \varphi_{\kappa(\epsilon)}(\epsilon,\alpha) \\ &\quad \times \int \Delta\alpha' p_{eq}(\epsilon,\alpha')\varphi_{\kappa(\epsilon)}(\epsilon,\alpha').\end{aligned} \tag{8.34}$$

For $t\to\infty$ only the eigenvalue $k(\epsilon)=0$ is retained; then this becomes, recalling $p_{eq}(\epsilon,\alpha')=p_{eq}(\epsilon,\alpha)$, $\varphi_0 = 1/\sqrt{G(\epsilon)}$, and $\int\Delta\alpha' = G(\epsilon)$,

$$\begin{aligned}\lim_{t\to\infty}\langle B_d^R(t)\rangle &= \int_0^\infty d\epsilon \int \Delta\alpha \langle\epsilon\alpha|B|\epsilon\alpha\rangle p_{eq}(\epsilon,\alpha) \\ &= \operatorname{Tr}\{\rho_{eq} B_d\} = \langle B_d\rangle.\end{aligned} \tag{8.35}$$

Also, for $\Delta B_d^R(t)$, defined analogous to (3.13) of the Kubo theory, we have

$$\lim_{t\to\infty}\langle \Delta B_d^R(t)\rangle = \lim_{t\to\infty}\operatorname{Tr}\{\rho_{eq}[B_d^R(t) - I\langle B_d\rangle]\} = 0. \tag{8.36}$$

Thus, the approach to equilibrium is well established.

We also consider the correlations. From (8.33) and (8.22b),

$$\begin{aligned}&\operatorname{Tr}\{\rho_{eq} B_d^R(t)A\} \\ &= \int_0^\infty d\epsilon \mathop{S}_{\kappa(\epsilon)} \exp[-\kappa(\epsilon)t] \int \Delta\alpha \langle\epsilon\alpha|B|\epsilon\alpha\rangle \varphi_{\kappa(\epsilon)}(\epsilon,\alpha) \int_0^\infty d\epsilon' \\ &\quad \times \int \Delta\alpha' p_{eq}(\epsilon',\alpha')\varphi_{\kappa(\epsilon')}(\epsilon'\alpha')\delta(\epsilon-\epsilon')\langle\epsilon'\alpha'|A|\epsilon'\alpha'\rangle.\end{aligned} \tag{8.37}$$

Now let $t\to\infty$. Again with $p_{eq}(\epsilon',\alpha')=p_{eq}(\epsilon',\alpha)$ and $\varphi_0 = 1/\sqrt{G(\epsilon)}$ we obtain

$$\begin{aligned}&\lim_{t\to\infty}\langle B_d^R(t)A\rangle \\ &= \int_0^\infty d\epsilon \int \Delta\alpha \langle\epsilon\alpha|B|\epsilon\alpha\rangle p_{eq}(\epsilon,\alpha) \int_0^\infty d\epsilon' \\ &\quad \times \int \Delta\alpha' \langle\epsilon'\alpha'|A|\epsilon'\alpha'\rangle \delta(\epsilon-\epsilon')\frac{1}{G(\epsilon)}.\end{aligned} \tag{8.38}$$

If the ensemble is microcanonical, then $p_{eq} = \delta(\epsilon-\epsilon')/G(\epsilon)$. So for that case the mixing property follows,

$$\lim_{t\to\infty}\langle B_d^R(t)A\rangle = \langle B\rangle\langle A\rangle. \tag{8.39}$$

This result can also be obtained from the Schrödinger form of the correlation function, using (8.14).

C. CONSEQUENCES FOR TRANSPORT FORMULAS

9. SUSCEPTIBILITY AND CONDUCTIVITY FOR CLASSICAL FREQUENCIES

We consider the Kubo form for the response function of the interacting system under consideration. We have the reduced function

$$\phi_{BA}(t) = \int_0^\beta d\beta' \operatorname{Tr}\{\rho_{eq}^0 \dot{A}^I(-i\hbar\beta') \lim_\lambda B^H(t)\}. \tag{9.1}$$

Here ρ_{eq}^0 is the canonical distribution $\exp(-\beta H^0)/\operatorname{Tr}\exp(-\beta H^0)$; $\dot{A}^I$ is the operator in the interaction picture

$$\dot{A}^I(t) = \exp(iH^0 t/\hbar)\dot{A}\exp(-iH^0 t/\hbar). \tag{9.2}$$

A problem is now that we cannot directly replace B^H by B_d^H and take the van Hove limit; this procedure would result in removing from the trace all terms $\langle\gamma|\dot{A}^I(-i\hbar\beta')|\gamma'\rangle$ with $\gamma\neq\gamma'$. But, the diagonal element $\langle\gamma|\dot{A}^I|\gamma\rangle = (1/i\hbar)\langle\gamma|[A^I, H^0]|\gamma\rangle = 0$. Suppose, however, that in (9.1) we first take the classical frequency limit (5.4). Then from stationarity,

$$\begin{aligned}&\langle\dot{A}^I(-i\hbar\beta')\lim_\lambda B^H(t)\rangle \\ &= \lim_\lambda \langle\dot{A}^H(0)B^H(t+i\hbar\beta')\rangle \\ &\approx \lim_\lambda \langle\dot{A}^H(0)\dot{B}^H(t)\rangle = \langle\dot{A}^S \lim_\lambda B^H(t)\rangle.\end{aligned} \tag{9.3}$$

Here $\dot{A}^S$ is the Schrödinger operator, for which the superscript will be dropped.[49] Substituting (9.3) into (9.1) we obtain

$$\phi_{BA}^{cl}(t) = \beta \operatorname{Tr}\{\rho_{eq}^0 \dot{A} \lim_\lambda B^H(t)\}, \tag{9.4}$$

where the superscript cl indicates that this is valid for classical frequencies $\hbar\omega \ll 1/\beta$. Similarly for the relaxation function

$$\Psi_{BA}^{cl}(t) = \beta \operatorname{Tr}\{\rho_{eq}^0 \Delta A \lim_\lambda \Delta B^H(t)\}. \tag{9.5}$$

9.1. The interaction form

We now set $B^H(t) - B_d^H(t)$ and use the result (6.8) of the preceding sections. We then obtain the following interaction forms

$$\phi^{\rm cl}_{BA}(t)=\beta\,{\rm Tr}\{\rho^0_{\rm eq}\dot{A}\exp(-\Lambda_d t)B\},\tag{9.6}$$

$$\Psi^{\rm cl}_{BA}(t)=\beta\,{\rm Tr}\{\rho^0_{\rm eq}\,\Delta A\exp(-\Lambda_d t)\,\Delta B\}.\tag{9.7}$$

By (3.6), (3.9), and (3.11) we find the corresponding expressions for χ and L. For example, from (3.6),

$$\chi^{\rm cl}_{BA}(i\omega)=\beta\int_0^\infty dt\exp(-i\omega t)\,{\rm Tr}\{\rho^0_{\rm eq}\dot{A}\exp(-\Lambda_d t)B\}.\tag{9.8}$$

9.2. The Schrödinger form

With the trace computed in the representation $|\gamma\rangle$, the result (9.5) also reads

$$\phi^{\rm cl}_{BA}(t)=\beta\int\Delta\gamma p^0_{\rm eq}(\gamma)\langle\gamma|\dot{A}|\gamma\rangle\langle\gamma|B^R_d(t)|\gamma\rangle.\tag{9.9}$$

The latter matrix element we take from (7.5) and we write $\dot{A}_\gamma=\langle\gamma|\dot{A}|\gamma\rangle$, etc. We then find

$$\phi^{\rm cl}_{BA}(t)=\beta\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)B_{\gamma'}\dot{A}_\gamma P(\gamma',t|\gamma).\tag{9.10}$$

Likewise for (9.7) and (9.8),

$$\Psi^{\rm cl}_{BA}(t)=\beta\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)\Delta B_{\gamma'}\Delta A_\gamma P(\gamma',t|\gamma),\tag{9.11}$$

$$\chi^{\rm cl}_{BA}(i\omega)=\beta\int_0^\infty dt\exp(-i\omega t)\int\int\Delta\gamma\Delta\gamma'\times p^0_{\rm eq}(\gamma)B_{\gamma'}\dot{A}_\gamma P(\gamma',t|\gamma).\tag{9.12}$$

Similarly for the conductivity

$$L^{\rm cl}_{BA}(i\omega)=\beta\int_0^\infty dt\exp(-i\omega t)\int\int\Delta\gamma\Delta\gamma'\times p^0_{\rm eq}(\gamma)\dot{B}_{\gamma'}\dot{A}_\gamma P(\gamma',t|\gamma).\tag{9.13}$$

For the d.c. electrical conductivity ($\dot{\mathbf{A}}=\dot{\mathbf{B}}=\mathbf{J}$) this result was also given by van Hove[50] though seemingly with one-particle results (as in Boltzmann theory, $|\gamma\rangle\to|\mathbf{k}\rangle$ = Bloch state) in mind.

We may also introduce a two-point probability by Bayes' rule

$$W_2(\gamma',t;\gamma,0)=P(\gamma',t|\gamma)p^0_{\rm eq}(\gamma).\tag{9.14}$$

Then we have the stochastic espressions

$$\chi^{\rm cl}_{BA}(i\omega)=\beta\int_0^\infty dt\exp(i\omega t)\int\int\Delta\gamma\Delta\gamma'\dot{A}_\gamma B_{\gamma'}W_2(\gamma',t;\gamma,0),\tag{9.15}$$

$$L^{\rm cl}_{BA}(i\omega)=\beta\int_0^\infty dt\exp(-i\omega t)\times\int\int\Delta\gamma\Delta\gamma'\dot{A}_\gamma\dot{B}_{\gamma'}W_2(\gamma',t;\gamma,0).\tag{9.16}$$

We refer to (9.10)—(9.16) as the Schrödinger form, since in this form the operators are fixed, while the time dependence is vested in the probability function. A summary of results is given in Table IV.

The connection between the Kubo form—based on the Heisenberg operators—and the Schrödinger form can also be shown directly. With $\rho^H=\rho^S(0)=\rho_{\rm eq}$ and $B^H(t)=U^\dagger(t)B^SU(t)$ we have quite generally, $\{|\varphi\rangle\}$ being an arbitrary complete set of states,

$$\begin{aligned}\langle B^H(t)A^H(0)\rangle&={\rm Tr}\{B^H(t)A^H(0)\rho^H\}={\rm Tr}\{B^H(t)A^S\rho_{\rm eq}\}\\&=\sum_{\varphi,\varphi',\varphi'',\varphi''',\varphi^{\rm iv}}\langle\varphi|U^\dagger|\varphi'\rangle\langle\varphi'|B^S|\varphi''\rangle\\&\quad\times\langle\varphi''|U|\varphi'''\rangle\langle\varphi'''|A^S|\varphi^{\rm iv}\rangle\langle\varphi^{\rm iv}|\rho_{\rm eq}|\varphi\rangle.\end{aligned}\tag{9.17}$$

We now assume that B^S and A^S are replaceable by diagonal operators, and we make an initial random phase assumption. The result is

$$\begin{aligned}\langle B^H(t)A^H(0)\rangle&=\sum_{\varphi\varphi'}|\langle\varphi|U^\dagger|\varphi'\rangle|^2\langle\varphi'|B^S|\varphi'\rangle\\&\quad\times\langle\varphi|A^S|\varphi\rangle\langle\varphi|\rho_{\rm eq}|\varphi\rangle\\&=\sum_{\varphi\varphi'}P(\varphi',t|\varphi)B^S_{\varphi'}A^S_\varphi p_{\rm eq}(\varphi).\end{aligned}\tag{9.18}$$

Next we introduce the interaction Hamiltonian (4.2), we choose $\{|\varphi\rangle\}=\{|\gamma\rangle\}$ and impose the van Hove limit. We then have $P(\varphi',t|\varphi)\to P(\gamma',t|\gamma)$, $p_{\rm eq}(\varphi)\to p^0_{\rm eq}(\gamma)$. This gives the previous result. The burden of the proof is, of course, to show that $P(\gamma',t|\gamma)$ satisfies the master equation, so that the formalism of Sec. 7 is still necessary in order to give the result the physical meaning set forth in this paper.

9.3. Justification of omission of nondiagonal parts

The rather qualitative argument by van Kampen[40] that macroscopic variables, displaying a rather slow time behavior, are mainly diagonal with the Hamiltonian, can readily be made explicit in the present context. We

TABLE IV. New forms (classical frequencies).

Function, coefficient	Interaction form		Schrödinger form	
$\phi^{\rm cl}_{BA}$	$\beta\,{\rm Tr}\{\rho^0_{\rm eq}\dot{A}\exp(-\Lambda_d t)B\}$	(9.6)	$\beta\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)B_{\gamma'}\dot{A}_\gamma P(\gamma',t\|\gamma)$	(9.10)
$\Psi^{\rm cl}_{BA}$	$\beta\,{\rm Tr}\{\rho^0_{\rm eq}\Delta A\exp(-\Lambda_d t)\Delta B\}$	(9.7)	$\beta\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)\Delta B_{\gamma'}\Delta A_\gamma P(\gamma',t\|\gamma)$	(9.11)
$\Phi^{\rm cl}_{BA}$	${\rm Tr}\{\rho^0_{\rm eq}\Delta A\exp(-\Lambda_d t)\Delta B\}$	(9.29)	$\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)\Delta B_{\gamma'}\Delta A_\gamma P(\gamma',t\|\gamma)$	
$\chi^{\rm cl}_{BA}$	$\beta\int_0^\infty dt\exp(-i\omega t){\rm Tr}\{\rho^0_{\rm eq}\dot{A}\exp(-\Lambda_d t)B\}$		$\beta\int dt\exp(-i\omega t)\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)B_{\gamma'}\dot{A}_\gamma P(\gamma',t\|\gamma)$	
$L^{\rm cl}_{BA}$	$\beta\int_0^\infty dt\exp(-i\omega t){\rm Tr}\{\rho^0_{\rm eq}\dot{A}\exp(-\Lambda_d t)\dot{B}\}$		$\beta\int dt\exp(-i\omega t)\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)\dot{B}_{\gamma'}\dot{A}_\gamma P(\gamma',t\|\gamma)$	
$\chi^{\rm cl}_{BA}-\chi^{\rm dc}_{BA}$	$-i\omega\beta\int_0^\infty dt\exp(-i\omega t){\rm Tr}\{\rho^0_{\rm eq}\Delta A\exp(-\Lambda_d t)\Delta B\}$		$-i\omega\beta\int_0^\infty dt\exp(-i\omega t)\int\int\Delta\gamma\Delta\gamma' p^0_{\rm eq}(\gamma)\Delta B_{\gamma'}\Delta A_\gamma P(\gamma',t\|\gamma)$	

consider the correlations $\langle \Delta\dot{B}(t)\Delta\dot{A}\rangle$. With $B = B_d + B_{nd}$, $A = A_d + A_{nd}$, one easily shows

$$\mathrm{Tr}\{\rho_{eq}\Delta\dot{B}(t)\Delta\dot{A}\} = \mathrm{Tr}\{\rho_{eq}\Delta\dot{B}_d(t)\Delta\dot{A}_d\} + \mathrm{Tr}\{\rho_{eq}\Delta\dot{B}_{nd}(t)\Delta\dot{A}_{nd}\}. \tag{9.19}$$

Thus, the Lemma (3.44) results in the following relation between diagonal and nondiagonal parts

$$\int_{-\infty}^{\infty} dt\exp(-i\omega t)[\mathrm{Tr}\{\rho_{eq}\Delta\dot{B}_{nd}(t)\Delta\dot{A}_{nd}\} - \exp(-\beta\hbar\omega)\,\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_{nd}\Delta\dot{B}_{nd}(t)\}] = [\exp(-\beta\hbar\omega)-1]\int_{-\infty}^{\infty} dt\exp(-i\omega t)\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_d\Delta\dot{B}_d(t)\}. \tag{9.20}$$

Consider now classical frequencies $\beta\hbar\omega \ll 1$. Expanding (9.20) we obtain

$$-\frac{1}{\hbar\omega}\int_{-\infty}^{\infty} dt\exp(-i\omega t)[\mathrm{Tr}\{\rho_{eq}\Delta\dot{B}_{nd}(t)\Delta\dot{A}_{nd}\} - \mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_{nd}\Delta\dot{B}_{nd}(t)\}] \approx \beta\int_{-\infty}^{\infty} dt\exp(-i\omega t)\,\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_d\Delta\dot{B}_d(t)\}. \tag{9.21}$$

But,

$$\Delta\dot{B}_{nd}(t)\Delta\dot{A}_{nd} - \Delta\dot{A}_{nd}\Delta\dot{B}_{nd}(t) = \Delta\dot{B}(t)\Delta\dot{A} - \Delta\dot{A}\Delta\dot{B}(t). \tag{9.22}$$

Thus the lhs of (9.21) is the Fourier transformed response function, cf. (3.47). So we have the result

$$\hat{\phi}^{cl}_{BA}(\omega) \approx \beta\int_{-\infty}^{\infty} dt\exp(-i\omega t)\,\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_d\Delta\dot{B}_d(t)\}, \tag{9.23}$$

or also,

$$\phi^{cl}_{BA}(t) \approx \beta\,\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_d\Delta\dot{B}_d(t)\}. \tag{9.24}$$

This result must remain valid in the van Hove limit:

$$\phi^{clR}_{BA}(t) \approx \beta\,\mathrm{Tr}\{\rho^0_{eq}\dot{A}B^R_d(t)\} \tag{9.25}$$

(we superscripted the response function with R for clarity, we omitted the Δ's since $\langle\dot{A}\rangle = 0$, and we used the fact that $\mathrm{Tr}\{\rho_{eq}\dot{A}_{nd}B^R_d(t)\} \equiv 0$). Equation (9.25) indicates that it suffices to know the diagonal part $B^R_d(t)$. Indeed, comparison with (9.4),

$$\phi^{clR}_{BA}(t) = \beta\,\mathrm{Tr}\{\rho^0_{eq}\dot{A}B^R(t)\}, \tag{9.26}$$

shows that for classical frequencies or for large times

$$B^R_{nd}(t) \ll B^R_d(t). \tag{9.27}$$

For quantum frequencies or for small times no such greatly simplifying statement can be made ("small" times means here $t \ll \hbar\beta$, but we still assume $t \gg \tau_t$ in accordance with the van Hove limit). From (9.20) one finds for $\beta\hbar\omega \gg 1$,

$$\mathrm{Tr}\{\rho_{eq}\Delta B_{nd}\Delta\dot{A}_{nd}\} \sim -\,\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_d\Delta B_d(t)\}. \tag{9.28}$$

Thus the nondiagonal contributions are no longer negligible. Results for quantum frequencies thus require a study of the nondiagonal reduced Heisenberg operators.

9.4. Classical fluctuation-dissipation theorems

For the relaxation function we found, cf. (9.5),

$$\Psi^{clR}_{BA}(t) = \beta\,\mathrm{Tr}\{\rho_{eq}\Delta\dot{A}_d\Delta B^R_d(t)\} = \beta\Phi^{clR}_{BA}(t), \tag{9.29}$$

i.e., Eq. (3.34) is confirmed for the reduced functions. May it be once more emphasized that the fluctuation–correlation function Φ^{clR}, in contrast to Φ^{cl}, shows clear-cut relaxation. Its Fourier spectrum exists,

$$S^{cl}_{BA}(\omega) = 2\int_{-\infty}^{\infty} dt\exp(-i\omega t)\Phi^{clR}_{BA}(t) = 2\int_0^{\infty} dt\exp(-i\omega t)\Phi^{clR}_{BA}(t) + 2\int_0^{\infty} dt\exp(i\omega t)\Phi^{clR}_{AB}(t) = 2\mathrm{Tr}\left\{\rho_{eq}\Delta A_d\frac{1}{\Lambda_d + i\omega}\Delta B_d\right\} + 2\mathrm{Tr}\left\{\rho_{eq}\Delta B_d\frac{1}{\Lambda_d - i\omega}\Delta A_d\right\}, \tag{9.30}$$

where we used the transposition property (next section). Further, from (9.29),

$$S^{cl}_{BA}(\omega) = 2kT\int_{-\infty}^{\infty} dt\exp(-i\omega t)\Psi^{clR}_{BA}(t), \tag{9.31}$$

which is the classical analog [$\mathcal{E}(\omega, T) = kT$] of (3.56). Thus the classical theorems analogous to (3.60) and (3.61) follow.

10. SYMMETRY PROPERTIES; THE ONSAGER-CASIMIR RELATIONS

The symmetry properties involving time and magnetic field reversal, given by Kubo (Ref. 2, Sec. 6) are in a much sharper way established in the present context. We consider the Schrödinger form for the classical fluctuation–correlation function

$$\Phi^{clR}_{BA}(t) = \int\int \Delta\gamma\Delta\gamma'\Delta B_{\gamma'}\Delta A_{\gamma}W_2(\gamma', t; \gamma, 0). \tag{10.1}$$

First, switching the integration variables, we also have

$$\Phi^{clR}_{BA}(t) = \int\int \Delta\gamma\Delta\gamma'\Delta B_{\gamma}\Delta A_{\gamma'}W_2(\gamma, t; \gamma', 0). \tag{10.2}$$

But

$$W_2(\gamma, t; \gamma', 0) = W_2(\gamma, 0; \gamma', -t) = W_2(\gamma', -t; \gamma, 0), \tag{10.3}$$

by stationarity and pair invariance. Comparing now (10.2), (10.3) with (10.1) we find the *transposition property*

$$\Phi^{clR}_{BA}(t, \mathbf{H}) = \Phi^{clR}_{AB}(-t, \mathbf{H}), \tag{10.4}$$

where the magnetic field has been added for comparison with what follows.

Secondly, we consider the time-reversal features stated in (7.14). The result was based on the time-reversal relation $U(-t) = U^{\dagger}(t)$, for which we assumed that the Hamiltonian is invariant, $H(-t) = H(t)$. However, in order that this be so, we must switch the magnetic field $\mathbf{H}$ along with t; thus $[U^{\dagger}(t)]^{\mathbf{H}} = [U(-t)]^{-\mathbf{H}}$; this leads to

$$P^{\mathbf{H}}(\gamma, t\,|\,\gamma') = P^{-\mathbf{H}}(\gamma', -t\,|\,\gamma). \tag{7.14)–(10.5}$$

Likewise, Eq. (7.13) must be amended, to read

$$P^{-\mathbf{H}}(\gamma, -t\,|\,\gamma') = \tilde{g}^{\mathbf{H}}(\gamma, -t;\ \gamma'). \qquad (7.13)-(10.6)$$

Combining these two equations, and using $P^{\mathbf{H}} \equiv g^{\mathbf{H}}$, we find that reciprocity (7.15) is still satisfied

$$g^{\mathbf{H}}(\gamma, t; \gamma') = \tilde{g}^{\mathbf{H}}(\gamma', -t;\ \gamma). \qquad (7.15)=(10.7)$$

Also, the mathematical identity (7.16) still holds, so that (7.16) and (7.19) are still valid, i.e., $P^{\mathbf{H}}$ is symmetrical:

$$P^{\mathbf{H}}(\gamma, t\,|\,\gamma') = P^{\mathbf{H}}(\gamma', t\,|\,\gamma). \qquad (7.19)=(10.8)$$

Finally, combining (10.5) and (10.8) we have

$$P^{\mathbf{H}}(\gamma, t\,|\,\gamma') = P^{-\mathbf{H}}(\gamma, -t\,|\,\gamma'), \qquad (10.9)$$

or also

$$W_2^{\mathbf{H}}(\gamma, t;\ \gamma', 0) = W_2^{-\mathbf{H}}(\gamma, -t;\ \gamma', 0). \qquad (10.10)$$

The time reversal further alters the Schrödinger operators A and B according to $\epsilon_A A$ and $\epsilon_B B$, with $\epsilon = \pm 1$, depending on the even $(+1)$ or odd (-1) character of the variables. Thus from (10.2) and (10.10) we have the *time-reversal property*,

$$\Phi_{BA}^{\mathrm{cl}R}(t, \mathbf{H}) = \epsilon_A \epsilon_B \Phi_{BA}^{\mathrm{cl}R}(-t, -\mathbf{H}). \qquad (10.11)$$

Combining (10.4) and (10.11) we also have

$$\Phi_{BA}^{\mathrm{cl}R}(t, \mathbf{H}) = \epsilon_A \epsilon_B \Phi_{AB}^{\mathrm{cl}R}(t, -\mathbf{H}). \qquad (10.12)$$

We note the microscopic nature of this proof: $\Phi^{\mathrm{cl}R}$ was linked directly to the Hamiltonian via $H \to U \to P \to W_2 \to \Phi$.

Since $\Psi^{\mathrm{cl}R}$ is proportional to $\Phi^{\mathrm{cl}R}$, the symmetry properties also hold for $\Psi^{\mathrm{cl}R}$. Further, the result (10.12) is immediately carried over to the Fourier–Laplace transforms $\chi(i\omega)$ and $L(i\omega)$. We thus obtain, restricting ourselves to classical frequencies

$$\chi_{BA}(i\omega, \mathbf{H}) = \epsilon_A \epsilon_B \chi_{AB}(i\omega, -\mathbf{H}) \qquad (10.13)$$

and, noticing $\epsilon_A \epsilon_B = \epsilon_A \epsilon_B$,

$$L_{BA}(i\omega, \mathbf{H}) = \epsilon_A \epsilon_B L_{AB}(i\omega, -\mathbf{H}). \qquad (10.14)$$

These are the Onsager[51]–Casimir[52] reciprocity relations.

ACKNOWLEDGMENTS

I am first of all indebted to R. J. J. Zijlstra and C. Th. J. Alkemade of the University of Utrecht, the Netherlands, for creating the opportunity for me to lecture on linear response theory in March 1977, in connection with which many ideas were developed and shaped. I would like to thank N. G. van Kampen for the arrangement of a colloquium and for sharing his criticism and profound reflections on this topic. Further I have benefited from discussions with M. Dresden, G. Vojta, and E. G. D. Cohen, and from discussions and a colloquium arranged by S. R. de Groot, A. J. Kox, and W. A. van Leeuwen at the University of Amsterdam. In Montreal I would like to thank my colleague A. Royer for comments with regard to the Liouville space.

APPENDIX A: DERIVATION OF THE MATRIX ELEMENTS IN VAN HOVE'S FORMALISM

The derivation of the matrix elements $\langle\gamma\,|\,B_d(t)\,|\,\gamma'\rangle$ (where B_d is an arbitrary diagonal operator) goes somewhat smoother than the derivation of $P(\gamma, t\,|\,\gamma')$ found in van Hove's paper, since there is more symmetry. Though projection operator techniques—see Appendix B—are faster, van Hove's method has the advantage that the computation is basically direct; we still note that our starting point is the Liouville equation which leads directly to (6.9) whose solutions we seek in the van Hove limit, while van Hove's treatment is based on the Schrödinger equation, an unnecessary complication. For this and other reasons we include this derivation here.

The essential feature of the development is the construction of a series expansion for $U(t)$ which is different from that in (6.10). For the present problem it is advantageous to split U directly into a diagonal part (U_d) and a nondiagonal part (U_{nd})—"diagonal" referring to the representation $\{|\gamma\rangle\}$ which are eigenstates of H^0. The main results needed are contained in two lemmas.

Lemma 1: The diagonal part $U_d(t) = \sum_{n=0}^{\infty} U_d^{(2n)}(t)$ involves all even orders of perturbation and is summable,

$$U_d = \exp(-iH^0 t/\hbar) \sum_{\gamma} |\gamma\rangle \exp[i(\lambda^2 t/\hbar)\Delta(\gamma) - (\lambda^2 t/\hbar)\Gamma(\gamma)]\langle\gamma|, \qquad (A1)$$

where

$$\Gamma(\gamma) = \pi \int \Delta\beta X(\beta, \gamma)\delta(\epsilon_\beta - \epsilon_\gamma), \qquad (A2)$$

$$\Delta(\gamma) = P \int \Delta\beta X(\beta, \gamma)/(\epsilon_\beta - \epsilon_\gamma), \qquad (A3)$$

P denoting the Cauchy principal value.

Proof: The evolution operator is written as $\sum_{k=0}^{\infty} U^{(k)}$ where $U^{(k)}$ for $k \geq 1$ follows by iteration from (6.9'),[41]

$$U^{(k)} = (\lambda/i\hbar)^k \int_0^t d\tau_k \int_0^{\tau_{k-1}} d\tau_{k-1} \cdots \int_0^{\tau_2} d\tau_1 \times \exp[-iH^0(t-\tau_k)/\hbar] V(\tau_k) \exp[-iH^0(\tau_k - \tau_{k-1})/\hbar] \times V(\tau_{k-1}) \cdots V(\tau_1) \exp[-iH^0 \tau_1/\hbar]. \qquad (A4)$$

The matrix elements are now

$$\begin{aligned}\langle\gamma\,|\,U^{(k)}\,|\,\gamma_0\rangle &= (\lambda/i\hbar)^k \sum_{\gamma_{k-1}\cdots\gamma_1} \int_0^t d\tau_k \int_0^{\tau_k} d\tau_{k-1} \cdots \int_0^{\tau_2} d\tau_1 \\ &\times \exp[-i\epsilon_\gamma(t-\tau_k)/\hbar]\langle\gamma\,|\,V(\tau_k)\,|\,\gamma_{k-1}\rangle \\ &\times \exp[-i\epsilon_{\gamma_{k-1}}(\tau_k - \tau_{k-1})/\hbar]\langle\gamma_{k-1}\,|\,V(\tau_{k-1})\,|\,\gamma_{k-2}\rangle \cdots \\ &\times \langle\gamma_2\,|\,V(\tau_2)\,|\,\gamma_1\rangle \exp[-i\epsilon_{\gamma_1}(\tau_2 - \tau_1)/\hbar] \\ &\times \langle\gamma_1\,|\,V(\tau_1)\,|\,\gamma_0\rangle \exp(-i\epsilon_{\gamma_0}\tau_1/\hbar). \end{aligned} \qquad (A5)$$

We picture this in Fig. 1, representing the time points by points, the interaction by arrows taking place at certain time points, and the states by intervals between the time points which indicate how long a given state lasts (we differ slightly with van Hove, who represents

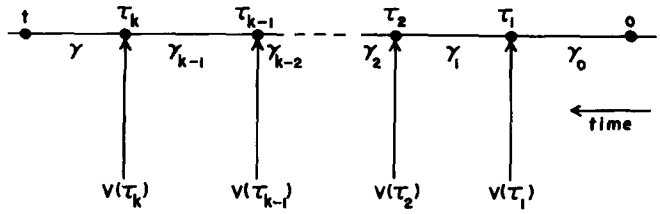


FIG. 1. Diagram for reading off the matrix element of the evolution $\langle\gamma|U^{(k)}|\gamma_0\rangle$.

the states by points). The structure of (A6) is readable directly from the figure.

We now consider the part $U^{(2n)}$ of the evolution operator; we write integrals for the state sums and set $V(t)\equiv V$,

$$\begin{aligned}\langle\gamma|U^{(2n)}|\gamma_0\rangle &= (\lambda/i\hbar)^{2n}\int_0^t d\tau_{2n}\int_0^{\tau_{2n}} d\tau_{2n-1}\cdots\int_0^{\tau_2} d\tau_1\int\Delta\gamma_{2n-1}\cdots\int\Delta\gamma_1\\ &\times\exp[-i\epsilon_\gamma(t-\tau_{2n})/\hbar]\langle\gamma|V|\gamma_{2n-1}\rangle\\ &\times\exp[-i\epsilon_{\gamma_{2n-1}}(\tau_{2n}-\tau_{2n-1})/\hbar]\langle\gamma_{2n-1}|V|\gamma_{2n-2}\rangle\cdots\\ &\times\langle\gamma_1|V|\gamma_0\rangle\exp(-i\epsilon_{\gamma_0}\tau_1/\hbar).\end{aligned}\tag{A6}$$

We especially consider the double transitions such as $\gamma\to\gamma_{2n-1}\to\gamma_{2n-2}$; the rule (6.13) is applied to the operator trio $V\{\ \}V$ as follows,

$$\begin{aligned}&\langle\gamma|V\{\ |\gamma_{2n-1}\rangle\exp[-i\epsilon_{\gamma_{2n-1}}(\tau_{2n}-\tau_{2n-1})/\hbar]\langle\gamma_{2n-1}|\{\ \}V|\gamma_{2n-2}\rangle\\ &=\delta_{\gamma,\gamma_{2n-2}}\exp[-i\epsilon_{\gamma_{2n-1}}(\tau_{2n}-\tau_{2n-1})/\hbar]X(\gamma_{2n-1},\gamma)\\ &\quad+Y\text{ part.}\end{aligned}\tag{A7}$$

Preceding and following this trio are exponentials, which combined with the exponential in (A7) and taking account of the Kronecker delta of (A7) can be written as

$$\begin{aligned}&\exp[-i\epsilon_\gamma(t-\tau_n)/\hbar]\exp[-i\epsilon_{\gamma_{2n-1}}(\tau_{2n}-\tau_{2n-1})/\hbar]\\ &\times\exp[-i\epsilon_{\gamma_{2n-2}}(\tau_{2n-1}-\tau_{2n-2})/\hbar]\\ &=\exp(-i\epsilon_\gamma t/\hbar)\exp[i(\epsilon_\gamma-\epsilon_{\gamma_{2n-1}})(\tau_{2n}-\tau_{2n-1})/\hbar]\\ &\times\exp(i\epsilon_\gamma\tau_{2n-2}/\hbar).\end{aligned}\tag{A8}$$

We now repeat the trio operation for other pairs, such as $\langle\gamma_{2n-2}|V\{\ \}V|\gamma_{2n-4}\rangle$, etc. There are a total of n such pairs, yielding the deltas $\delta_{\gamma,\gamma_{2n-2}}$, $\delta_{\gamma_{2n-2},\gamma_{2n-4}}$, $\delta_{\gamma_2,\gamma_0}$. We thus have "saw tooth" transitions in which each second neighbor is equal to the original state; $\gamma\to\gamma_{2n-1}\to\gamma\to\gamma_{2n-3}\to\gamma$, etc. The states $\gamma_{2n-1},\gamma_{2n-3},\ldots,\gamma_1$ are called free states. Moreover, leaving off the Y parts, we pick out the diagonal matrix element $\gamma=\gamma_0$, as follows from integration over all Kronecker deltas

$$\begin{aligned}&\int\Delta\gamma_{2n-2}\cdots\int\Delta\gamma_2\,\delta_{\gamma,\gamma_{2n-2}}\delta_{\gamma_{2n-2},\gamma_{2n-4}}\cdots\delta_{\gamma_2,\gamma_0}\\ &=\int d\gamma_{2n-2}\cdots\int d\gamma_2\,\delta(\gamma-\gamma_{2n-2})\delta(\gamma_{2n-2}-\gamma_{2n-4})\cdots\\ &\quad\times\delta(\gamma_2-\gamma_0)/Z(\gamma_0)\\ &=\delta(\gamma-\gamma_0)/Z(\gamma_0)=\delta_{\gamma,\gamma_0}.\end{aligned}\tag{A9}$$

The result (A6) is now complete, except that we must still integrate over the free states $\int\Delta\gamma_{2n-1}\int\Delta\gamma_{2n-3}\cdots\times\int\Delta\gamma_1$. We relabel these variables, viz., $\gamma_{2n-1}=\beta_n$, $\gamma_{2n-3}=\beta_{n-1},\ldots,\gamma_1=\beta_1$. From (A7)—(A9) the result is then

$$\begin{aligned}&\langle\gamma|U_d^{(2n)}|\gamma_0\rangle\\ &=\delta_{\gamma,\gamma_0}\exp(-i\epsilon_\gamma t/\hbar)(\lambda/i\hbar)^{2n}\int_0^t d\tau_{2n}\\ &\quad\times\int_0^{\tau_{2n}}d\tau_{2n-1}\cdots\int_0^{\tau_2}d\tau_1\int\Delta\beta_n\cdots\int\Delta\beta_1\\ &\quad\times\exp[i(\epsilon_\gamma-\epsilon_{\beta_n})(\tau_{2n}-\tau_{2n-1})/\hbar]\\ &\quad\times\exp[i(\epsilon_\gamma-\epsilon_{\beta_{n-1}})(\tau_{2n-2}-\tau_{2n-3})/\hbar]\cdots\\ &\quad\times\exp[i(\epsilon_\gamma-\epsilon_{\beta_1})(\tau_2-\tau_1)/\hbar]X(\beta_n,\gamma)X(\beta_{n-1},\gamma)\cdots X(\beta_1,\gamma)\end{aligned}\tag{A10}$$

Introducing the time variables $t-\tau_{2n}=\theta_{2n}$, $\tau_{2n}-\tau_{2n-1}=\theta_{2n-1},\ldots,\tau_2-\tau_1=\theta_1$, the time integral becomes

$$\begin{aligned}I(t)=&\int_0^t d\theta_{2n}\int_0^{t-\theta_{2n}}d\theta_{2n-1}\cdots\int_0^{t-(\theta_{2n}+\cdots+\theta_2)}d\theta_1\\ &\times\exp[i(\epsilon_\gamma-\epsilon_{\beta_n})\theta_{2n-1}/\hbar]\exp[i(\epsilon_\gamma-\epsilon_{\beta_{n-1}})\theta_{2n-3}/\hbar]\cdots\\ &\times\exp[i(\epsilon_\gamma-\epsilon_{\beta_1})\theta_1/\hbar].\end{aligned}\tag{A11}$$

For the Laplace transform this yields by the convolution theorem

$$\begin{aligned}\hat I(s)&=\frac{1}{s^n}\prod_{k=1}^n\int_0^\infty\exp(-s\theta)\exp[i(\epsilon_\gamma-\epsilon_{\beta_k})\theta/\hbar]\,d\theta\\ &=\frac{1}{s^{n+1}}\prod_{k=1}^n\frac{1}{s-(i/\hbar)(\epsilon_\gamma-\epsilon_{\beta_k})}\\ &\approx\frac{1}{s^{n+1}}\prod_{k=1}^n\left(\frac{\hbar i/\rho}{\epsilon_\gamma-\epsilon_{\beta_k}}+\pi\hbar\delta(\epsilon_\gamma-\epsilon_{\beta_k})\right),\end{aligned}\tag{A12}$$

since $|s|\sim 1/t\ll\delta\epsilon/\hbar$. So also

$$\begin{aligned}&\int\Delta\beta_n\cdots\int\Delta\beta_1X(\beta_n,\gamma)\cdots X(\beta_1,\gamma)\hat I(s)\\ &\quad=(1/s^{n+1})[\hbar\Gamma(\gamma)-\hbar i\Delta(\gamma)]^n;\end{aligned}\tag{A13}$$

whence finally by inversion,

$$\langle\gamma|U_d^{(2n)}|\gamma_0\rangle=\delta_{\gamma,\gamma_0}\exp(-i\epsilon_\gamma t/\hbar)\left(\frac{-\lambda^2 t}{\hbar}\right)^n\frac{[\Gamma(\gamma)-i\Delta(\gamma)]^n}{n!}.\tag{A14}$$

It is also clear that the odd operators $U^{(2n+1)}$ do not yield diagonal contributions since the saw-tooth scheme fails; the extra time integral causes the magnitude to go with $\lambda\to 0$. Summing the even order operators one has therefore,

$$\langle\gamma|U_d|\gamma_0\rangle=\delta_{\gamma,\gamma_0}\exp(-i\epsilon_\gamma t/\hbar)\exp[-(\lambda^2 t/\hbar)[\Gamma(\gamma)-i\Delta(\gamma)]]\tag{A15}$$

which is the diagonal form of (A1).

Lemma 2: The nondiagonal part of the evolution operator U_{nd} is expressible in the diagonal part U_d by a series; thus the total U is

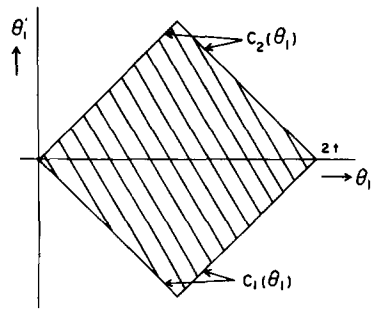


FIG. 2. Integration of the time integrals.

$$U = U_d + \sum_{n=1}^{\infty} (-i\lambda/\hbar)^n \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \cdots \int_0^{\tau_2} d\tau_1 \times \{U_d(t-\tau_n) V U_d(\tau_n - \tau_{n-1}) V \cdots V U_d(\tau_1)\}_{\langle nd \rangle}; \quad \text{(A16)}$$

the subscript $\langle nd \rangle$ means that in the curly bracket the intermediary states and the initial state must all be kept different from the final state. Van Hove gives this result without proof. We will show that this result is easily obtained with the projector operator method, see the next appendix.

The evaluation of the matrix elements is now straightforward. The diagonal part of U yields, cf. (A15),

$$\langle \gamma | U_d^\dagger B_d U_d | \gamma' \rangle = \delta_{\gamma\gamma'} \langle \gamma | B_d | \gamma \rangle \exp[-2\lambda^2 t \Gamma(\gamma)/\hbar^2]. \quad \text{(A17)}$$

The cross terms of U_d and U_{nd} obviously yield no diagonal terms. For the nondiagonal part we exemplify the procedure by looking at the first term,

$$\langle \gamma | U_{nd}^{(1)\dagger} B_d U_{nd}^{(1)} | \gamma' \rangle = \frac{\lambda^2}{\hbar^2} \int_0^t d\tau_1 \int_0^t d\tau_1' u_{d\gamma}^*(\tau_1') \times \langle \gamma | \underbrace{V\{U_d^\dagger(t-\tau_1')}_{(nd)} B_d \underbrace{U_d(t-\tau_1)\}V}_{(nd)} | \gamma' \rangle u_{d\gamma'}(\tau_1). \quad \text{(A18)}$$

Here, $u_{d\gamma}$ are the eigenvalues of U_d,

$$U_d(t) | \gamma \rangle = u_{d\gamma}(t) | \gamma \rangle = \exp[-i\epsilon_\gamma t/\hbar + (\lambda^2 t/\hbar)[i\Delta(\gamma) - \Gamma(\gamma)]] | \gamma \rangle. \quad \text{(A19)}$$

(The requirement that there are no diagonal transitions is automatic for this first step since V is nondiagonal.)

We apply van Hove's functional rule for the operator $\{\ \}$ in (A18). Then,

$$\text{matrix el.} = \frac{\lambda^2}{\hbar^2} \int_0^t d\tau_1 \int_0^t d\tau_1' u_{d\gamma}^*(\tau_1') u_{d\gamma}(\tau_1) \times \{\delta_{\gamma\gamma'} \int \Delta\beta u_{d\beta}^*(t-\tau_1') u_{d\beta}(t-\tau_1) \langle \beta | B_d | \beta \rangle X(\beta, \gamma) + \int \Delta\beta u_{d\beta}^*(t-\tau_1') u_{d\beta}(t-\tau_1) \langle \beta | B_d | \beta \rangle Y(\beta; \gamma, \gamma')\}. \quad \text{(A20)}$$

To evaluate the term with X we set $\tau_1 - \tau_1' = \theta_1'$, $\tau_1 + \tau_1' = \theta_1$. The area of integration is indicated in Fig. 2.

We denote by C_2 the path along the upper two sides and by C_1 the path over the lower two sides. The result then is

$$\text{matrix el.} = \frac{\lambda^2}{2\hbar^2} \delta_{\gamma\gamma'} \int \Delta\beta \langle \beta | B_d | \beta \rangle X(\beta, \gamma) \int_0^{2t} d\theta_1 \int_{C_1(\theta_1)}^{C_2(\theta_1)} d\theta_1' \times \exp\{(i/\hbar)\{\epsilon_\beta - \epsilon_\gamma - \lambda^2[\Delta(\beta) + \Delta(\gamma)]\}\theta_1'\} \times \exp\{(\lambda^2/\hbar)[\Gamma(\beta) - \Gamma(\gamma)]\theta_1\} \exp[-2(\lambda^2 t/\hbar)\Gamma(\beta)]. \quad \text{(A21)}$$

To evaluate the integral over $d\theta_1'$, we note the asymptotic form

$$\int_{-\delta T}^{\delta T} d\theta_1' \exp[i(\alpha - \alpha')\theta_1'] \approx 2\pi\delta(\alpha - \alpha'), \quad \delta\alpha\delta T \gg 1. \quad \text{(A22)}$$

For the majority of values along the curves $C_1(\theta_1)$ and $C_2(\theta_1)$, this relation holds. In the resulting delta functions, we can set $\lambda^2 \to 0$. Thus the result for this part is

$$(\text{matrix el.})_{\text{due to } X} = \frac{\pi\lambda^2}{\hbar} \delta_{\gamma\gamma'} \int \Delta\beta \langle \beta | B_d | \beta \rangle X(\beta, \gamma) \int_0^{2t} d\theta_1 \times \delta(\epsilon_\beta - \epsilon_\gamma) \exp[-(\lambda^2/\hbar)\Gamma(\beta)(2t - \theta_1)] \times \exp[-(\lambda^2/\hbar)\Gamma(\gamma)\theta_1]. \quad \text{(A23)}$$

The part due to Y can be evaluated similarly, employing Eqs. (6.18) for the integrals. The result is found to be of order λ^2, hence vanishing.

The complete story can now be evaluated in the same manner. For simplicity we write in what follows $\bar{\lambda} = \lambda/\sqrt{\hbar}$. The general matrix element is

$$\langle \gamma | U_{nd}^{(n)\dagger} B_d U_{nd}^{(m)} | \gamma' \rangle = (\bar{\lambda}/\sqrt{\hbar}i)^m (\bar{\lambda} i/\sqrt{\hbar})^n \int_0^t d\tau_n \cdots \int_0^{\tau_2} d\tau_1 \times \int_0^t d\tau_n' \cdots \int_0^{\tau_2'} d\tau_1' \int \Delta\gamma_{n-1} \cdots \int \Delta\gamma_1 \int \Delta\gamma_{n-1}' \cdots \int \Delta\gamma_1' \times u_{d\gamma}^*(\tau_1')[\langle \gamma | V | \gamma_1' \rangle u_{d\gamma_1}^*(\tau_2' - \tau_1') \langle \gamma_1' | V | \gamma_2' \rangle \cdots \times u_{d\gamma_{n-1}}^*(\tau_n' - \tau_{n-1}') \langle \gamma_{n-1}' | V\{U_d^\dagger(t-\tau_n')\}]_{\langle nd \rangle} B_d \times [U_d(t-\tau_m)\}V | \gamma_{m-1} \rangle u_{d\gamma_{m-1}}(\tau_m - \tau_{m-1}) \langle \gamma_{m-1} | V | \gamma_{m-2} \rangle \cdots \times \langle \gamma_1 | V | \gamma' \rangle]_{\langle nd \rangle} u_{d\gamma'}(\tau_1). \quad \text{(A24)}$$

We apply van Hove's functional rule first to the operators $V\{\ \}V$ occurring in the middle. Next we apply this rule to the expression between the two V's lying symmetrically outside the middle part, and so on. Thus we establish bridges between all pairs γ_i and γ_i'. No restrictions are placed on the intermediate states β_i which arise from the functional rule; however, we note that the diagonal transitions are always absent due to the structure of the kernels X [$X(\beta_n, \beta_{n-1})$ exists only for $\beta_n \neq \beta_{n-1}$]. In case $n = m$, the above procedure will exhaust all terms. The time integrals can be handled with a double Laplace transform, or, in a more elementary way, as in the above procedure, i.e., we make the

change in variables $\tau_i - \tau_i' = \theta_i'$, $\tau_i + \tau_i' = \theta_i$. In finding the new limits, we noted that the first change in variables $(\tau_n, \tau_n' \to \theta_n, \theta_n')$ resulted in limits $0 \leq \theta_n \leq 2t$ while for θ_n' we can set $-\delta T \leq \theta_n' \leq \delta T$, where δT, though asymptotically large for the integrand, is negligible compared to θ_n. Thus one readily sees that the integration area on θ_{n-1}, θ_{n-1}' is again as in Fig. 2, but with $0 \leq \theta_{n-1} \lesssim \theta_n$, while for θ_{n-1}' we can again set $-\delta T \leq \theta_{n-1}' \leq \delta T$; and so on. The result is easily found to be

$$\begin{aligned}
&\langle\gamma|\, U_{nl}^{(n)\dagger} B_d U_{nl}^{(n)} |\gamma'\rangle \\
&= (\pi\bar{\lambda}^2)^n \delta_{\gamma\gamma'} \int_0^{2t} d\theta_n \int_0^{\theta_n} d\theta_{n-1} \cdots \int_0^{\theta_2} d\theta_1 \\
&\times \int \Delta\beta_n \cdots \int \Delta\beta_1 \delta(\epsilon_{\beta_n} - \epsilon_{\beta_{n-1}}) \cdots \\
&\times \delta(\epsilon_{\beta_1} - \epsilon_\gamma) X(\beta_n, \beta_{n-1}) \cdots X(\beta_1, \gamma) \\
&\times \exp[-\bar{\lambda}^2 \Gamma(\beta_n)(2t - \theta_n)] \exp[-\bar{\lambda}^2 \Gamma(\beta_{n-1})(\theta_n - \theta_{n-1})] \cdots \\
&\times \exp[-\bar{\lambda}^2 \Gamma(\beta_1)(\theta_2 - \theta_1)] \exp[-\bar{\lambda}^2 \Gamma(\gamma)\theta_1] \langle\beta_n|\, B_d\, |\beta_n\rangle.
\end{aligned} \tag{A25}$$

If we compute in a like fashion the matrix elements for $n \neq m$, we must, in addition to the above, account for $(n - m)$ remaining integrals. The result is found to be of order $\lambda^{|n-m|}$, i.e., vanishing. Hence, summing (A25) for all orders n, we obtain the complete result, as given in the text Eq. (7.3). If we still affect the change in variables $\theta_i = 2\vartheta_i$, the time ordering is even more symmetric $(0 \leq \vartheta_1 \leq \vartheta_2, \ldots, 0 \leq \vartheta_n \leq t)$ and the formulas are in appearance similar to those of van Hove.

APPENDIX B: PROJECTION OPERATOR SOLUTION FOR THE EVOLUTION OPERATOR *U(t)*

The objective of finding expressions for the diagonal and nondiagonal parts of the evolution operator is most expediently accomplished by using Zwanzig's projector operator method.[53] Since, to our knowledge, this scheme has not been applied to the evolution in state space (Zwanzig's paper is mainly concerned with the Green's operator in Liouville space), we will do this here.[54]

We consider the differential equation (6.9) for $U(t)$ and we split into diagonal and nondiagonal parts. We note the general rule

$$\begin{aligned}
AB &= \{(A_d + A_{nd})(B_d + B_{nd})\}_d + \{(A_d + A_{nd})(B_d + B_{nd})\}_{nd} \\
&= (A_d B_d)_{(d)} + (A_d B_{nd})_{(nd)} + (A_{nd} B_d)_{(nd)} \\
&\quad + (A_{nd} B_{nd})_d + (A_{nd} B_{nd})_{nd}.
\end{aligned} \tag{B1}$$

The subscripts in parentheses are automatically fulfilled and can be omitted. Noting that H^0 has no nondiagonal terms and λV has no diagonal terms, Eq. (6.9) splits as follows:

$$i\hbar \frac{\partial U_d}{\partial t} = H^0 U_d + (\lambda V U_{nd})_d = H^0 U_d + P\lambda V U_{nd}, \tag{B2}$$

$$\begin{aligned}
i\hbar \frac{\partial U_{nd}}{\partial t} &= H^0 U_{nd} + \lambda V U_d + (\lambda V U_{nd})_{nd} \\
&= H^0 U_{nd} + \lambda V U_d + (1 - P)\lambda V U_{nd},
\end{aligned} \tag{B3}$$

where we revert to Zwanzig's notation, $A_d = PA$, $A_{nd} = (1 - P)A$; the reader is further reminded that the projection operator P or $1 - P$ acts on everything to its right, except when otherwise indicated by brackets.

The Green's operator of (B3) satisfies

$$\frac{\partial \mathcal{G}(t, t')}{\partial t} - \frac{1}{i\hbar}[H^0 \mathcal{G} + (1 - P)\lambda V \mathcal{G}] = I\delta(t - t'), \tag{B4}$$

where I is the unit operator in state space. The solution is

$$\mathcal{G}(t, t') = I \exp\{1/i\hbar)(t - t')[H^0 + (1 - P)\lambda V]\}\, u(t - t') \tag{B5}$$

The formal solution of (B3) becomes with the initial condition $U_{nd}(0) = 0$, and noting $\mathcal{G}(t, t') = \mathcal{G}(t - t', 0) \equiv \mathcal{G}(t - t')$,

$$\begin{aligned}
U_{nd} &= \frac{1}{i\hbar} \int_0^t dt' \mathcal{G}(t - t')\lambda V U_d(t') \\
&= \frac{1}{i\hbar} \int_0^t dt' \mathcal{G}(t')\lambda V U_d(t - t').
\end{aligned} \tag{B6}$$

When this is substituted into (B2) the result is

$$i\hbar \frac{\partial U_d}{\partial t} = H^0 U_d + \frac{\lambda^2}{i\hbar} \int_0^t dt' PV \mathcal{G}(t') V U_d(t - t'). \tag{B7}$$

This result is still exact. We now consider the weak coupling limit of small λ. Then for $\mathcal{G}(t')$ we write $\mathcal{G}^0(t')$ where

$$\mathcal{G}^0(t') = I \exp(-it'H^0/\hbar) u(t'). \tag{B8}$$

Then (B7) becomes approximately

$$i\hbar \frac{\partial U_d}{\partial t} = H^0 U_d + \frac{\lambda^2}{i\hbar} \int_0^t dt' PV \exp(-it'H^0/\hbar) V U_d(t - t'). \tag{B9}$$

We note the integral is of order $\lambda^2 t$.

Now let $W = \exp(iH^0 t/\hbar) U_d$. Then for W we have the integrodifferential equation

$$\begin{aligned}
\frac{\partial W}{\partial t} &= -\frac{\lambda^2}{\hbar^2} \exp(iH^0 t/\hbar) \int_0^t dt' \\
&\quad \times PV \exp(-it'H^0/\hbar) V \exp[-i(t - t')H^0/\hbar] W(t - t').
\end{aligned} \tag{B10}$$

We use van Hove's functional rule in the form (6.13′). Noticing that P does not have to act on the diagonal factors $\exp[-i(t - t')H^0/\hbar]$ and $W(t - t')$, we have

$$\begin{aligned}
&PV \exp(-it'H^0/\hbar) V \\
&= \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma|\, \exp(-it'\epsilon_{\gamma'}/\hbar) X(\gamma', \gamma);
\end{aligned} \tag{B11}$$

hence also,

$$\begin{aligned}
&\exp(iH^0 t/\hbar) P[V \exp(-it'H^0/\hbar) V] \exp[-i(t - t')H^0/\hbar] \\
&= \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma|\, \exp[-it'(\epsilon_{\gamma'} - \epsilon_\gamma)/\hbar] X(\gamma', \gamma).
\end{aligned} \tag{B12}$$

This is substituted into (B10). Taking the Laplace transform $\hat{W}(s)$ we obtain by convolution

$$s\hat{W}(s) - W(0) = -\frac{\lambda^2}{\hbar^2} \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma|\, \frac{X(\gamma', \gamma)}{s + i(\epsilon_{\gamma'} - \epsilon_\gamma)/\hbar} \hat{W}(s). \tag{B13}$$

Here $W(0) = U_d(0) = 1$. We now consider asymptotic times, or small $|s|$, so that

$$\frac{1}{s + i(\epsilon_{\gamma'} - \epsilon_\gamma)/\hbar} \approx \hbar \left[\frac{-i\mathcal{P}}{\epsilon_{\gamma'} - \epsilon_\gamma} + \pi\delta(\epsilon_{\gamma'} - \epsilon_\gamma) \right]. \tag{B14}$$

With the definitions (A2) and (A3), Eq. (B13) yields

$$\hat{W}(s) = \{ s + \frac{\lambda^2}{\hbar} \sum_\gamma |\gamma\rangle\langle\gamma| [\Gamma(\gamma) - i\Delta(\gamma)] \}^{-1}. \tag{B15}$$

The inverse transformation gives

$$W(t) = \exp\left\{ -\frac{\lambda^2 t}{\hbar} \sum_\gamma |\gamma\rangle\langle\gamma| [\Gamma(\gamma) - i\Delta(\gamma)] \right\}. \tag{B16}$$

Hence for U_d,

$$U_d = \exp\left\{ \frac{-it}{\hbar} H^0 - \frac{\lambda^2 t}{\hbar} \sum_\gamma |\gamma\rangle\langle\gamma| [\Gamma(\gamma) - i\Delta(\gamma)] \right\}. \tag{B17}$$

By series expansion one easily verifies that

$$\exp\{ -\sum_\gamma |\gamma\rangle\langle\gamma| F(\gamma) \} = \sum_\gamma |\gamma\rangle\langle\gamma| \exp\{ -F(\gamma) \}. \tag{B18}$$

Thus (B17) is in accord with (A1). We thus obtained van Hove's diagonal part in an easier way.

For the purpose of the nondiagonal part, see (B6), we may not approximate $g(t)$ by $g^0(t)$; we must maintain the complete result which is of order $(\lambda t)^n$ (all orders $n \geq 1$). We note that the van Hove limit cannot be carried through in $U_{nd}(t)$ itself, though terms of order $(\lambda^2 t)^n$ are readily extracted from the matrix elements $\langle\gamma| B_d(t) |\gamma'\rangle$, see Appendix A.

In the result (B6) we take the Laplace transform,

$$\hat{U}_{nd}(s) = \frac{1}{\hbar i} \frac{1}{s + (i/\hbar)[H^0 + (1-P)\lambda V]} \lambda V \hat{U}_d(s). \tag{B19}$$

For the computation of the matrix elements we will, as before, seek a convolution series in $U_d(t)$. Therefore, we rewrite (B19) as follows, using (B17)

$$\hat{U}_{nd}(s) = \frac{1}{\hbar i} \{ s + (i/\hbar)H^0 + (\lambda^2/\hbar)\phi - [(1/\hbar i)(1-P)\lambda V + (\lambda^2/\hbar)\phi] \}^{-1} \lambda V \hat{U}_d(s), \tag{B20}$$

where

$$\phi = \sum_\gamma |\gamma\rangle\langle\gamma| [\Gamma(\gamma) - i\Delta(\gamma)], \tag{B21}$$

$$\hat{U}_d(s) = \frac{1}{s + (i/\hbar)H^0 + (\lambda^2/\hbar)\phi}. \tag{B22}$$

Equation (B20) is now expanded in a perturbation series, see, e.g., Fano,[25]

$$\hat{U}_{nd}(s) = \hat{U}_d(s) \sum_{n=0}^{\infty} \left\{ \left[\frac{\lambda}{\hbar i}(1-P)V + \frac{\lambda^2}{\hbar}\phi \right] \hat{U}_d(s) \right\}^n \frac{\lambda}{\hbar i} V \hat{U}_d(s). \tag{B23}$$

In the factor in [ ], the term in λ^2 can be neglected versus the term in λ; in front of the last factor V we set $(1-P)$ for symmetry reasons. The result is then

$$\hat{U}_{nd}(s) = \hat{U}_d(s) \sum_{n=1}^{\infty} \left\{ \frac{\lambda}{\hbar i}(1-P)V \hat{U}_d(s) \right\}^n. \tag{B24}$$

The inverse transform is easily shown to be

$$\begin{aligned} U_{nd}(t) &= \sum_{n=1}^{\infty} \left(\frac{\lambda}{\hbar i}\right)^n \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \cdots \\ &\times \int_0^{\tau_2} d\tau_1 U_d(t-\tau_n)(1-P)VU_d(\tau_n - \tau_{n-1}) \\ &\times \cdots (1-P)VU_d(\tau_2 - \tau_1)(1-P)VU_d(\tau_1) \\ &= \sum_{n=1}^{\infty} \left(\frac{\lambda}{\hbar i}\right)^n \int_0^t d\tau_n \int_0^{\tau_n} d\tau_{n-1} \cdots \int_0^{\tau_2} d\tau_1 U_d(t-\tau_n) \\ &\times \{ VU_d(\tau_n - \tau_{n-1}) \{ V \cdots \{ VU_d(\tau_2 - \tau_1) \\ &\times \underbrace{\{ VU_d(\tau_1) \}_{nd} \cdots \}_{nd}}_{n\text{-fold}}. \end{aligned} \tag{B25}$$

The n-fold nd requirement defines the symbol $\langle nd \rangle$ of (A16); the full result is hereby proven.

[1] R. Kubo, Can. J. Phys. **34**, 1274 (1956).
[2] R. Kubo, J. Phys. Soc. Jpn. 12, 570 (1957).
[3] M. S. Green, J. Chem. Phys. 20, 1281 (1952); 22, 398 (1954).
[4] J. G. Kirkwood, J. Chem. Phys. 14, 180 (1946).
[5] H. B. Callen and T. A. Welton, Phys. Rev. 83, 34 (1951).
[6] R. F. Greene and H. B. Callen, Phys. Rev. 88, 1387 (1952).
[7] H. Nyquist, Phys. Rev. 32, 110 (1928).
[8] E. de Haas-Lorentz, *Die Brownsche Bewegung und einige anverwandte Erscheinungen* (Vieweg, Braunschweig, Germany).
[9] E. Verboven, Physica 26, 1091 (1960).
[10] S. Fujita and R. Abe, J. Math. Phys. 3, 350 (1962).
[11] N. G. van Kampen, Physica Norvegica 5, 279 (1971).
[12] G. V. Chester and A. Thellung, Proc. Phys. Soc. 73, 745 (1960).
[13] S. F. Edwards, Phil. Mag. 3, 33, 1020 (1958).
[14] D. A. Greenwood, Proc. Phys. Soc. 71, 585 (1958).
[15] M. Lax, Phys. Rev. 109, 1921 (1958).
[16] L. van Hove, Physica 21, 517 (1955).
[17] L. van Hove, in *Les gaz neutres et ionisés*, edited by C. de Witt. Ecole d'été de physique théorique, Les Houches (1955) (Hermann, Paris, 1960), p. 149.
[18] K. M. van Vliet, "Linear Response Theory Revisited. II: The Master Equation Approach," to be submitted to J. Math. Phys.
[19] R. Kubo, in *Lectures in Theoretical Physics*, Vol. I, edited by W. E. Britten and L. G. Durham, Boulder, Colorado (1958) (Interscience, New York, 1959).
[20] R. M. Mazo, *Statistical Mechanics of Transport Processes* (Pergamon, New York, 1967), Chap. 10.
[21] E. Montroll, in *Lectures in Theoretical Physics*, Vol. III, edited by W. E. Britten, B. W. Downs, and J. Downs, Boulder, Colorado, 1960 (Interscience, New York, 1961).
[22] M. E. van Valkenburg, *Network Analysis* (Prentice-Hall, Englewood Cliffs, New Jersey, 1955).
[23] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Dover, New York, 1960), footnote p. 4.
[24] When necessary we use the superscripts S for Schrödinger and and H for Heisenberg operators. However, in general an unbracketed operator is a Schrödinger operator, $A \equiv A^S = A^H(0)$ and a bracketed operator is a Heisenberg operator, $A(t) \equiv A^H(t)$, unless otherwise indicated.
[25] U. Fano, Rev. Mod. Phys. 29, 74 (1957); also in *Lectures on the Many-Body Problem*, Vol. 2, edited by E. R. Caianiello (Academic, New York, 1964).
[26] If $\dot{\mathbf{B}}$ is a vector flux, and A is a vector, then $\phi_{\mathbf{BA}}$ is a tensor. The extensions are straightforward.

[27]O. E. Lanford, III and J. L. Lebowitz, *Lecture Notes in Physics*, Vol. 38 (Springer, New York, 1974), p. 144.
[28]S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics* (North-Holland, Amsterdam; Interscience, New York, 1962).
[29]J. W. van Velsen, "On linear response theory and area preserving mappings," Ph. D. thesis, University of Utrecht (1976).
[30]Of course, one can also shift the operators, assuming $\mathrm{Tr}\rho_{eq}A = 0$. However, this does often not befit the physical picture (e. g., the mean magnetization in thermal equilibrium need not be zero). More important, this assumption detracts from the basic tenet that the transport coefficients are linked with the fluctuations in the system, as in the *fluctuation*—dissipation theorem.
[31]In particular (3.24) shows that there are no steady equilibrium fluxes, $\lim_{t\to\infty}\langle\dot{A}(t)\rangle = 0$, if this limit exists. It may not exist, however, in Kubo's original theory, though it fares better in the reinterpreted theory, see Sec. 8.
[32]The difficulties alluded to are exemplified here in that (3.22) when integrated over t leads to (3.30), but (3.22′) does not. This stems from the contributions involving $\langle A\rangle$ and $\langle B\rangle$. For example, the contribution $\mathrm{Tr}\{\rho_{eq}\dot{A}(-t')\langle B\rangle\}$ contained in the integrand based on (3.22′) is zero according to (3.24). However, integrating the trace term by term one obtains $-\langle A(-t')\rangle\langle B\rangle \neq 0$. These pitfalls are avoided in the present approach, since the convergence properties for $t\to\infty$ of the integrand based on (3.22) are better. [We also note that Kubo's original paper (Ref. 2) is in this respect in error, in that one cannot go from op. cit. (3.11b) to op. cit (4.5), though the latter is correct and in agreement with our (3.30).]
[33]S. Fujita, Physica 27, 1161 (1969); *Introduction to Nonequilibrium Statistical Mechanics* (Saunders, Philadelphia, 1968).
[34]K. M. van Vliet and J. R. Fassett, in *Fluctuation Phenomena in Solids*, edited by R. E. Burgess (Academic, New York, 1965), p. 267.
[35]N. G. van Kampen, Symposium in Commemoration of Ohm's Law, held at Cologne (May 1974), Proceedings in press.
[36]Of course, the method of solution is not in itself important. Thus, one can conceive of models which are soluble without perturbation theory, such as a collection of harmonic oscillators, consisting of n small masses and one big mass, coupled by springs, cf. R. J. Rubin, J. Am. Chem. Soc. 90, 5061 (1968) and J. Math. Phys. 1, 309 (1960). The motion of the big mass is essentially randomized by the coupling to the background motion of the small masses, as is clear from the generalized Langevin equation which results.
[37]L. van Hove, Physica 23, 441 (1957).
[38]F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionelle* Paris (Gauthier-Villars, Budapest, Akademiai Kiadó, 1965).
[39]R. V. Churchill, *Operational Mathematics* (McGraw-Hill, New York, 1958), p. 174.
[40]N. G. van Kampen, in *Fundamental Aspects of Statistical Mechanics. I* (North-Holland, Amsterdam, 1962), p. 173.
[41]A. Messiah, *Quantum Mechanics. II* (North-Holland, Amsterdam, 1962).
[42]Van Hove writes $\delta(\gamma-\gamma')$ instead of $\delta_{\gamma\gamma'}$ and refers to (6.13) as the diagonal-singularity property of large systems. In our opinion (6.13) is just a convenient rule; the fact that Y does not contribute is a direct consequence of weak interaction, rather than of a built-in matrix property.
[43]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Vol. I (McGraw-Hill, New York, 1953).
[44]We assume that there are no boundary terms and that any initial condition is included in $Q(\gamma)$ as an impulse $\delta(t)f(\gamma,0)$.
[45]For simplicity, we assume here that there is no magnetic field; for refinements, see Sec. 10.
[46]See E. H. Lieb and D. C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966).
[47]Since $p_{eq}(\epsilon\alpha)$ is independent of α, we could also have defined the scalar product in van Kampen's form [N. G. van Kampen, in *Fluctuation Phenomena in Solids*, edited by R. E. Burgess (Academic, New York, 1965), p. 139.]

$$(f,g)^{\epsilon} = \int \Delta\alpha f(\epsilon,\alpha)g(\epsilon,\alpha)/p(\epsilon\alpha).$$

This form is mandatory for the macroscopic master equation which is not self-adjoint.
[48]In Dirac notation $\{D|C\} = \{C,D\} = \mathrm{Tr}D^{\dagger}C$.
[49]We note that the classical frequency limit is not the same as the "classical limit" [as is suggested in several papers (Ref. 2, 20)]. In particular, A and B in the rhs of (9.3) are still quantum operators. It is important to recognize that we did not bluntly set $\hbar\to 0$; rather it suffices to impose that the combination $\hbar\beta/t\to 0$ or that $\hbar\beta\omega\to 0$.
[50]L. van Hove, in *Fundamental Aspects of Statistical Mechanics*. I (North-Holland, Amsterdam, 1962), p. 157.
[51]L. S. Onsager, Phys. Rev. 37, 405 (1937); 38, 265 (1938).
[52]H. B. G. Casimir, Rev. Mod. Phys. 17, 343 (1945).
[53]R. W. Zwanzig, in *Lectures in Theoretical Physics*, Vol. III, edited by W. E. Britten, B. W. Downs, and J. Downs, Boulder, Colorado (1960) (Interscience, New York, 1961), pp. 106–41.
[54]A more detailed derivation and discussion will be published in Can. J. Phys.

A no-go theorem for polarization structure[a)]

Gary R. Goldstein

Department of Physics, Tufts University, Medford, Massachusetts 02155

Michael J. Moravcsik

Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403
(Received 11 April 1977)

It would be extremely advantageous, from the point of view of the planning of experiments and of the testing of theoretical models by such experiments, if a representation of the reaction matrix could be found such that the differential cross section is given by the absolute value squared of a single amplitude, and the simplest type of polarization experiments are expressed in terms of functions involving only two amplitudes each. It is shown that, within the framework of a set of plausible and very weak constraints, such a representation does not exist.

I. INTRODUCTION

Polarization experiments constitute a wealth of information about particle reactions in high energy, nuclear, and atomic physics, and in fact provide indispensibly crucial data in the determination of the reaction matrix. As a result, much attention was devoted during the last two decades to the formulation of representations for the reaction matrices such that the relationship of amplitudes and experimental observables is simple, thus facilitating the planning of experiments and the testing of theoretical models. Since in a given theoretical model it is relatively easy to calculate any set of amplitudes one chooses to use for the reaction matrix, the representations should be constructed primarily with the compatibility with experiments in mind.

In terms of these experimental realities, priorities develop according to the degree of ease of the conventional experimental techniques, and not in terms of the information content of the particular experiment. Thus, for example, measuring differential cross sections (i.e., the observable in which no particle is polarized) is naturally always the first order of business, in spite of the fact that perhaps of all observables, this one has the least information content from the point of view of the exploration of the reaction matrix. Beyond the differential cross section, in general, the fewer particles that need to be simultaneously polarized in the experiment, the easier is the experiment judged to be.

On the other hand, even apart from the lack of information content of the differential cross section, it is expressed in virtually all representations of the reaction matrix by a cumbersome form involving the sum of the squares of the absolute values of *all* of the reaction amplitudes. This makes it additionally useless in the experimental determination of the reaction amplitudes.

For this reason it appears particularly attractive to try to find a representation for the reaction matrix in which the amplitudes are chosen as follows. The differential cross section is expressed in terms of the absolute value squared of one single amplitude. The remaining amplitudes are chosen in such a way that a set of simple experiments can be found which completely determines the reaction amplitudes through a set of simple relations between the experimental observables and the amplitudes, so that each of these relations involves only two amplitudes. With the help of such relations, the algorithm of actually obtaining the values of the amplitudes from experimental measurements would be very easy and free of unnecessary ambiguities, and the propagation of the uncertainties could also be traced easily from the measurements to the amplitudes.

The purpose of this note is to demonstrate that under a set of very plausible and weakly constraining assumtions, such a dream-representation cannot be realized. Thus we will have to continue to consider the differential cross section as an observable which either should not be included at all in a set aimed at determining completely the reaction amplitudes, or must remain a member of a set which is cumbersome and possibly wasteful of experiments in determining the amplitudes. Hence our result provides a certain kind of a lower limit for the simplicity of the set of experiments that can in fact determine the reaction amplitudes.

II. THE THEOREM AND ITS PROOF

The theorem we want to demonstrate can be stated as follows: It is impossible to construct a representation of a reaction matrix so that the differential cross section is given by the absolute value squared of a single amplitude, and all the other observables in a set of observables which fully determine the amplitudes are given by functions of two amplitudes each.

In the proof of the theorem, we will need to demonstrate the truth of this statement only for the two types of reactions $0+0\to 0+s$ (where s is a boson) and $0+s_1 \to 0+s_2$, where s_1 and s_2 are fermions, and the s's denote both the name of the particles and the values of their spins which however can be chosen arbitrarily. To demonstrate the theorem for these two cases is sufficient because of the factorization theorem[1] which tells us that more general reactions can always be synthesized nondynamically out of reactions of the above types. This factorization theorem states that for all purposes which are independent of the particular dynamics governing the particle reaction, a reaction

[a)]Work in part supported by the U.S. Energy Research and Development Administration.

 0022-2488/78/1906-1371$1.00

containing several particles with nonzero spin can be synthesized out of simpler reactions each of which contains only a fewer number of such particles with nonzero spins. Since in this paper we deal only with the relationship of the reaction matrix and the experimental observables, which relationship is indeed independent of dynamics, the proof of our theorem needs to be given only for such simpler, irreducible reactions, and it will then hold automatically also for composite reactions.

Another simplification in the proof is brought about by the freedom we have in our choice of using, as a basis for comparison with the hypothetical new representation, any of the "traditional" representations of the reaction matrix. Such a comparison is needed in formulating the constraint among the bilinear combinations ("bicoms") of the amplitudes in the new representation. Utilizing this freedom of choice, we will make use of the optimal type representation[2] which enormously simplifies the relationship between observables and bicoms and hence makes the proof very much more transparent. Whether the optimal type representation does or does not conveniently match directly the experiments that can be simply carried out in the laboratory is of no relevance from this point of view, since if the proof of the theorem holds with an optimal representation, it also holds with any other representation that is linearly related to the optimal one, i.e., with any of the presently known and used representations. As will be seen, all of our arguments will be dimensional so linear transformations of the observables will not change our conclusions.

It might be worth remarking that if we restricted ourselves to representations of the reaction matrix which are *linearly* connected to the "traditional" representations, the theorem and its proof would be trivial. We want, however, to include a much broader class of representations of the reaction matrix, given by Eq. (7) below, thus making the theorem significant and at the same time the proof more subtle.

III. PROOF OF THEOREM

First let us consider $0+0 \to 0+s$. Without imposing any symmetry, other than Lorentz or rotation invariance, this reaction is described by $n=2s+1$ amplitudes, which we will call $f_1, f_2, \ldots, f_n$, in an optimal representation.[2] In such a representation all the observables are related to the bicoms by a matrix consisting only of submatrices along the diagonal, which, for irreducible reactions are of size 1×1 or 2×2. For reactions of the above type only 1×1's appear (i.e., the matrix is diagonal).

There are a number of nonlinear relations among the n^2 bicoms, that reduce the number of *independent* bicoms to $2n-1$. These constraints are of the following forms. Let

$$R_{ij}\equiv \mathrm{Re} f_i f_j^*, \quad I_{ij}\equiv \mathrm{Im} f_i f_j^*, \tag{1}$$

where $ij=1,2,\ldots,n$ and the R's and I's are now observables; and let

$$\Sigma_{ij}\equiv R_{ij}^2+I_{ij}^2. \tag{2}$$

Then the constraints are of the form[3]

$$\Sigma_{ij}=R_{ii}R_{jj}, \tag{3}$$

and

$$\arctan\frac{I_{ij}}{R_{ij}}+\arctan\frac{I_{jk}}{R_{jk}}+\arctan\frac{I_{ki}}{R_{ki}}=0. \tag{4}$$

There are $\binom{n}{2}$ relations of type (3) and $\binom{n}{3}$ of type (4), though they are not all independent. Using the type (3) relations we can obtain an important relation,

$$\binom{n-1}{3}\sum_{i=1}^{n}R_{ii}=\sum_{i<j<k=1}^{n}\frac{\Sigma_{ij}\Sigma_{jk}+\Sigma_{ij}\Sigma_{jk}+\Sigma_{ik}\Sigma_{jk}}{(\Sigma_{ij}\Sigma_{jk}\Sigma_{ik})^{1/2}}, \tag{5}$$

where the sum on the right involves $\binom{n}{3}$ terms. The sum on the left is proportional to the differential cross section, denoted by σ.

Let us now define a new set of amplitudes, $g_1, g_2, \ldots, g_n$, and assume that

$$\sigma=|g_1|^2=\sum_{i=1}^{n}R_{ii}. \tag{6}$$

Then the other observables will be real or imaginary parts of functions of the set $\{g\}$,

$$R_{ij}=h_{ij}(\{g\}), \quad I_{ij}=k_{ij}(\{g\}). \tag{7}$$

We now assume that the h_{ij}'s and k_{ij}'s are only functions of two g_i's each, i.e.,

$$h_{ij}(\{g\})=h_{ij}(g_i,g_j), \quad k_{ij}(\{g\})=k_{ij}(g_i,g_j),$$

and that these functions have no essential singularities in the g_i's.

The dimensions of the amplitudes $\{f\}$ are fixed by relation (6) and so it follows from (1) that observables R_{ij} and I_{ij} all have the dimensions of σ. Then the h_{ij}'s and k_{ij}'s all have the dimensions of σ as well. The amplitude g_1 has the same dimension as any of the f_i's ($\sigma^{1/2}$). Without loss of generality, therefore, we can assume that all the g_i's have the same dimension, $g_i\sim\sigma^{1/2}$. The functional forms of the h_{ij}'s and k_{ij}'s are thereby restricted to be real rational functions in the g_i's, viz.

$$h_{ij}(g_i,g_j)=\sum_{n,p=-\infty}^{\infty}c_{ij}^{(np)}g_i^n g_j^{*p}+\text{c.c.}$$

and (8)

$$k_{ij}(g_i,g_j)=\sum_{n,p=-\infty}^{\infty}d_{ij}^{(np)}g_i^n g_j^{*p}+\text{c.c.},$$

where $c_{ij}^{(np)}$ and $d_{ij}^{(np)}$ are arbitrary complex coefficients of dimension L^{2-n-p} (L meaning "length"). We require that

$$h_{ij}=h_{ji} \quad \text{and} \quad k_{ij}=-k_{ji}. \tag{9}$$

Next we turn to the constraint (4). Consider the simplest special case of the functions in (8). We have

$$\begin{aligned} h_{ij}&=L^{2-m_{ij}-q_{ij}}g_i^{m_{ij}}g_j^{*q_{ij}}+\text{c.c.},\\ k_{ij}&=iL^{2-n_{ij}-r_{ij}}g_i^{n_{ij}}g_j^{*r_{ij}}+\text{c.c.} \end{aligned} \tag{10}$$

Then

$$\arctan\frac{k_{ij}}{h_{ij}}=\arctan\left(\frac{|g_i|^{n_{ij}}|g_j|^{r_{ij}}\sin[n_{ij}\alpha_i-r_{ij}\alpha_j]}{|g_i|^{m_{ij}}|g_j|^{q_{ij}}\cos[m_{ij}\alpha_i-q_{ij}\alpha_j]}\times L^{(m_i f^{*q_{ij}})-(n_i f^{*r_{ij}})}\right), \tag{11}$$

where $g_i=|g_i|\exp(i\alpha_i)$.

Because of the symmetry requirements (9) on the h's and k's, we must have

$$m_{ij}=q_{ij} \quad \text{and} \quad n_{ij}=r_{ij}. \tag{12}$$

As a result, we have

$$\arctan\frac{k_{ij}}{h_{ij}}=\arctan\Big((|g_i|\ |g_j|)^{n_{ij}-m_{ij}}L^{2m_{ij}-2n_{ij}} \times\frac{\sin[n_{ij}(\alpha_i-\alpha_j)]}{\cos[m_{ij}(\alpha_i-\alpha_j)]}\Big). \tag{13}$$

We now substitute (13) into (4), for some particular values of (i,j,k). Equation (4) must hold for any values of the amplitude. In particular, suppose that $|g_i|$ is varied. Then the invariance of (4) requires that the partial derivative with respect to $|g_i|$ of the left-hand side vanishes. Hence we have

$$0=\frac{\partial}{\partial|g_i|}\arctan\frac{k_{ij}}{h_{ij}}+\frac{\partial}{\partial|g_i|}\arctan\frac{k_{ki}}{h_{ki}} =\frac{h_{ij}k_{ij}(n_{ij}-m_{ij})}{|g_i|(h_{ij}^2+k_{ij}^2)}+\frac{h_{ki}k_{ki}(n_{ki}-m_{ki})}{|g_i|(h_{ki}^2+k_{ki}^2)}. \tag{14}$$

This must hold for any values of g_j or g_k. Furthermore, the numerator and the denominator of the first term are not of the same order in g_j and cancellation of the dependence on g_j cannot occur in general. Therefore, we must have

$$n_{ij}=m_{ij}, \tag{15}$$

and similarly

$$n_{ki}=m_{ki}. \tag{16}$$

With these relations we substitute back into (13) and obtain from (4)

$$n_{ij}(\alpha_i-\alpha_j)+n_{jk}(\alpha_j-\alpha_k)+n_{ki}(\alpha_k-\alpha_i)=0. \tag{17}$$

In order for this to hold for all $\alpha_{i,j,k}$ we must have

$$n_{ij}=n_{jk}=n_{ki}, \tag{18}$$

so h_{ij} and k_{ij} are simply real and imaginary parts of $(g_ig_j^*)^n$.

Then we consider Eq. (5). It is apparent that the right-hand side is a symmetric function of all of the g_i's. But the left-hand side is just $\binom{n-1}{3}|g_1|^2$. Thus Eqs. (4) and (5) *cannot* be satisfied by any nonzero function of the form specified in (10).

Returning now to the study of constraints given in Eq. (4), we consider the next simplest case of the functions in (8); namely,

$$\begin{aligned}h_{ij}&=L^{2-m_{ij}-q_{ij}}g_i^{m_{ij}}g_j^{*q_{ij}} +L^{2-m'_{ij}-q'_{ij}}g_i^{m'_{ij}}g_j^{*q'_{ij}}+(i\rightleftarrows j)+\text{c.c.},\\ k_{ij}&=iL^{2-n_{ij}-r_{ij}}g_i^{n_{ij}}g_j^{*r_{ij}} +iL^{2-n'_{ij}-r'_{ij}}g_i^{n'_{ij}}g_j^{*r'_{ij}}+(i\rightleftarrows j)+\text{c.c.}\end{aligned} \tag{19}$$

Now again requiring that the partial derivative with respect to $|g_i|$ of Eq. (4) be zero, leads, as before, to the conclusion that

$$\frac{\partial}{\partial|g_i|}\arctan\frac{k_{ij}}{h_{ij}},$$

is independent of $|g_j|$. So the numerator and denominator must be the same order in $|g_i|$. The powers of $|g_j|$ in the denominator are

$$\begin{aligned}&2q,\ q+m,\ q+q',\ q+m',\ 2m,\ m+q',\ m+m',\\ &2q',\ q'+m',\ 2m',\ 2r,\ r+m,\ r+r',\ r+n',\ 2n,\\ &n+r',\ n+n',\ 2r',\ r'+n',\ 2n'.\end{aligned}$$

The numerator can be written as

$$h_{ij}\frac{\partial}{\partial|g_i|}k_{ij}-k_{ij}\frac{\partial}{\partial|g_i|}h_{ij}, \tag{20}$$

and therefore the powers of $|g_j|$ in the numerator will be

$$\begin{aligned}&q+r,\ q+n,\ q+r',\ q+n',\ m+n,\\ &m+r',\ m+n',\ q'+r',\ q'+n',\ m'+n'.\end{aligned} \tag{21}$$

We will now show that for these two sets to be the same, we must have pairwise equality between the powers that appear in h_{ij} and the powers that appear in k_{ij}. For this purpose, consider two sets of integers:

$$\{a\}=\overbrace{a_1a_2\cdots a_i}^{i}\ \overbrace{a_{i+1}\cdots a_p}^{p-i}\ \overbrace{a_{p+1}\cdots a_{n-k}}^{n-k-p}\ \overbrace{a_{n-k+1}\cdots a_k}^{k}$$

$$\{b\}=\underbrace{b_1b_2\cdots b_j}_{j}\ \underbrace{b_{j+1}\cdots b_q}_{q-j}\ \underbrace{b_{q+1}\cdots b_{n-l}}_{n-l-q}\ \underbrace{b_{n-l+1}\cdots b_n}_{l}$$

overlapping

e e o o

Between the two sets, we have an overlapping set of $n-k-i=n-l-j$ quantities; that is,

$$\{a_{i+1}\cdots a_pa_{p+1}\cdots a_{n-k}\}=\{b_{j+1}\cdots b_qb_{q+1}\cdots b_{n-l}\},$$

as sets. Which a is equal to which b is unknown and unimportant. We see from the above that $k+i=j+l$.

In each of these sets, some quantities are even (e) and some are odd (o), as indicated above. In particular, we see that among the nonoverlapping quantities i of the a's and j of the b's are even, and k of the a's and l of the b's are odd.

We now form the table of sums of these quantities; i.e., a table that contains at a given place the sums of the two quantities which label the row and the column of that particular place. The result is given in Table I. We see that the table is symmetric around the main diagonal. The parity of the sums is indicated in the table.

The sums are now ordered into two sets. Set I is $\{a_r+a_s,\ b_r+b_s\}$ while Set II is $\{a_r+b_s\}$. The table also shows which squares are in Set I, and which are in Set II, and which are in both sets.

Now we explore the conditions under which Set I is identical to Set II. For this we can ignore the overlap set which is already common between the two sets.

In each of the two remaining, nonoverlapping sets, we count up the number of even and the number of odd quantities, and then demand that the two numbers of the evens be the same, and the two numbers of the odds also be the same. We have

TABLE I. Table of the sums of the powers of g's in the numerator and denominator of Eq. (14). The letters e and o denote even and odd. The I and II refer to the two sum sets explained in the text.

		i e $a_1 \cdots a_i$	k o $a_n \cdots a_{n-k+1}$	j e $b_1 \cdots b_j$	l o $b_{j+1} \cdots b_{n-l+1}$	$p-i$ e $b_{j+1} \cdots b_q$ $a_{i+1} \cdots a_p$	$n-k-p$ o $b_{q+1} \cdots b_{n-l}$ $a_{p+1} \cdots a_{n-k}$	
e	$a_1 \cdots a_i$	e	o	e	o	e	o	i
o	$a_n \cdots a_{n-k+1}$	o	e	o	e	o	e	k
e	$b_1 \cdots b_j$	e	o	e	o	e	o	j
o	$b_{j+1} \cdots b_{n-l+1}$	o	e	o	e	o	e	l
e	$b_{j+1} a_{i+1} \cdots b_q a_p$	e	o	e	o	e	o	$p-i$
o	$b_{q+1} a_{p+1} \cdots b_{n-l} a_{n-k}$	o	e	o	e	o	e	$n-k-p$

I

II

I+II

Set I: Even: $\frac{1}{2}i(i+1)+\frac{1}{2}k(k+1)+\frac{1}{2}j(j+1)+\frac{1}{2}l(l+1)$,

Odd: $ik+jl$;

Set II: Even: $ij+kl$,

Odd: $il+jk$.

In addition, we also have $i+k=j+l$, as mentioned above.

From the two numbers of odds we have immediately $i=j$, and then from $i+k=j+l$ we get $k=l$. Finally, from the two numbers of evens we obtain $i+k=0$, and since i and k are positive integers, we must have $i=k=0$, and hence also $j=l=0$. We see, therefore, that the nonoverlapping sets are empty, and *all* a's and b's are in the overlapping set.

As a result, if the numerator and the denominator in Eqs. (20) and (21) are to have the same powers, the exponents must satisfy

$$q_{ij}=r_{ij}, \quad m_{ij}=n_{ij}, \quad q'_{ij}=r'_{ij}, \quad m'_{ij}=n'_{ij}. \tag{22}$$

Any permutation of the right-hand sides produces the same result.

Substituting these values of the exponents back into the derivative relations given by Eq. (14), we obtain independence of $|g_j|$ only if all the exponents are equal. Thus we have reduced this case to the simplest case.

The reasoning used in this second simplest case can be applied to the functions of the general form in (9) to obtain the same result, that only terms in which all exponents are equal survive the constraints (4), and then (5) cannot be satisfied.

This completes the proof for the reaction $0+0 \to 0+s$.

Now we consider reactions $0+s_1 \to 0+s_2$. In the optimal representation the matrix connecting the observables to the bicoms contains both 1×1 and 2×2 submatrices along the diagonal.[2]

The single polarization measurements can be constructed out of the 1×1 terms only, as can be seen in Table I of Ref. 1. It can also be seen from this table, combined with the rules given in Ref. 4, that the observables belonging to the 1×1's suffice to determine the amplitudes completely. Therefore, we need only consider the observables belonging to the 1×1 submatrices, because if a set of observables is complete in one representation, it is complete in all other representations. Hence these reactions are in no way different from the preceding reactions, and again no solution of the desired form can be obtained.

In conclusion, a representation of the amplitudes, in which the differential cross section completely determines the magnitude of a single amplitude, along with the constraint that single polarization observables be functions of pairs of amplitudes, can not be constructed. The impossibility of such a representation is a result of the strong restrictions imposed on any representation by the nonlinear constraint equations [(3), (4), and (5)] among the observables, as expressed in the optimal choice of those observables.

[1]P. L. Csonka, M. J. Moravcsik, and M. D. Scadron, Ann. Phys. (N. Y.) **41**, 1 (1967).
[2]G. R. Goldstein and M. J. Moravcsik, Ann. Phys. (N. Y.) **98**, 128 (1976).
[3]G. R. Goldstein, J. F. Owens III, J. P. Rutherfoord, and M. J. Moravcsik, Nucl. Phys. B **80**, 164 (1974).
[4]G. R. Goldstein, M. J. Moravcsik, and D. Bregman, Nuovo Cimento Lett. **11**, 137 (1974).

On the charged Kerr-Tomimatsu-Sato family of solutions

Masatoshi Yamazaki

Department of Physics, Kanazawa University, Kanazawa 920, Japan
(Received 10 November 1977)

The charged Kerr–Tomimatsu–Sato family of solutions with arbitrary integer distortion parameter δ for gravitational fields of spinning masses is presented. The Bonnor–Misra–Pandey–Srivastava–Tripathi–Wang family of solutions is also referred to.

PACS numbers: 04.20.Jb

1. CHARGED SOLUTIONS

Of interest recently has been the problem of finding the exact solutions of axisymmetric Einstein—Maxwell field equations.[1-10] The purpose of the present paper is to present the charged Kerr—Tomimatsu—Sato family of spinning mass solutions for arbitrary integer distortion parameter δ. It was shown that the Kerr[11] and the three Tomimatsu—Sato[12] spinning mass solutions can be written in a closed form[13] with arbitrary positive integer distortion parameter δ and that by direct substitution this closed form does satisfy[14] the equations on the first order intermediate integrals of the second order Ernst equations.[15] We have not yet succeeded in showing directly that this closed form[13] satisfies the Ernst equation.[15]

Now we have four parameters, i.e., the mass m, the angular momentum $J=m^2q\sigma=m\alpha$, the electric charge $e^2=m^2|\lambda|^2$, and the distortion parameter δ.

The family of solutions is of the form

$$ds^2=f^{-1}[e^{2\gamma}(dz^2+d\rho^2)+\rho^2\,d\phi^2]-f(dt-\omega\,d\phi)^2, \tag{1}$$

with

$$f=A/B, \tag{2}$$

$$\omega=-2mqbC/A, \tag{3}$$

$$\exp 2\gamma=A/p^{2\delta}(a-b)^{\delta^2}, \tag{4}$$

$$A=F(\delta^2), \tag{5}$$

$$B=A+(2/\sigma)H+\{1+(1/\sigma^2)\}G, \tag{6}$$

and

$$C=\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}a^{1-r'}b^{r-1}\{-pxg(\delta,r,r')-\tfrac{1}{2}\{\sigma+(1/\sigma)\}h(\delta,r,r')\}F(\delta^2-r), \tag{7}$$

and of the form

$$A_4=\lambda(\sigma H+G)/\sigma^2B, \tag{8}$$

$$A_3'=-\lambda I/\sigma B=-\lambda\Omega/2, \tag{9}$$

$$\frac{\partial A_3}{\partial x}=\omega\frac{\partial A_4}{\partial x}+\frac{B\kappa b}{A}\frac{\partial A_3'}{\partial y},\quad \frac{\partial A_3}{\partial y}=\omega\frac{\partial A_4}{\partial y}+\frac{B\kappa a}{A}\frac{\partial A_3'}{\partial x}, \tag{10}$$

and

$$A_3=\omega A_4+(\lambda\kappa/2\sigma^2)\{\sigma Q+R-(\delta/pq)\},$$

[see Eqs. (21) and (22) for Q and R], (11)

where A_3 and A_4 are the ϕ and t component of the electromagnetic 4-potential, respectively. We shall follow the notation of Ref. 13 unless otherwise noted. We define

$$\sigma^2+|\lambda|^2=1,\quad \alpha=mq\sigma,\quad \text{and } e=m|\lambda|. \tag{12}$$

We choose the unit of length as $\kappa=mp\sigma/\delta$, i.e.,

$$\rho=\kappa(x^2-1)^{1/2}(1-y^2)^{1/2}\quad\text{and}\quad z=\kappa xy. \tag{13}$$

The complex scalar electromagnetic potential[4] $\Phi=A_4+iA_3'$ and the twist potential Ω are

$$|\Phi|^2=|\lambda|^2G/\sigma^2B\quad\text{and}\quad\Omega=2I/\sigma B. \tag{14}$$

When $e=0(|\lambda|=0$ and $\sigma=1)$ the metric given in Eqs. (1)—(7) satisfies the vacuum Einstein equations.[13]

2. DERIVATION OF SOLUTIONS

It is straightforward to obtain Eqs. (2), (5), (6), (8), (9), and (14) from the procedure Ernst[4] has formulated. We shall derive Eqs. (3), (4), (7), (10), and (11) in this section. From Eq. (11) of Ref. 4 we obtain

$$\sigma\frac{\partial\omega}{\partial x}=\frac{2b\kappa}{A^2}\left[\left\{(A+2G)+\left(\frac{1}{\sigma}+\sigma\right)H\right\}\frac{\partial I}{\partial y}-I\frac{\partial}{\partial y}\left\{(A+2G)+\left(\frac{1}{\sigma}+\sigma\right)H\right\}\right] \tag{15}$$

and

$$\sigma\frac{\partial\omega}{\partial y}=\frac{2a\kappa}{A^2}\left[\left\{(A+2G)+\left(\frac{1}{\sigma}+\sigma\right)H\right\}\frac{\partial I}{\partial x}-I\frac{\partial}{\partial x}\left\{(A+2G)+\left(\frac{1}{\sigma}+\sigma\right)H\right\}\right]. \tag{16}$$

Now we have the following relations:

$$\frac{2a}{A^2}\left\{I\frac{\partial}{\partial x}(A+2G)-(A+2G)\frac{\partial}{\partial x}I\right\}=\frac{\partial Q}{\partial y}, \tag{17}$$

$$\frac{2b}{A^2}\left\{I\frac{\partial}{\partial y}(A+2G)-(A+2G)\frac{\partial}{\partial y}I\right\}=\frac{\partial Q}{\partial x}, \tag{18}$$

$$\frac{4a}{A^2}\left(I\frac{\partial H}{\partial x}-H\frac{\partial I}{\partial x}\right)=\frac{\partial R}{\partial y}, \tag{19}$$

$$\frac{4b}{A^2}\left(I\frac{\partial H}{\partial y}-H\frac{\partial I}{\partial y}\right)=\frac{\partial R}{\partial x}, \tag{20}$$

$$Q=-\frac{2px}{pqA}\,\delta\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}q^2b^ra^{1-r'}g(\delta,r,r')F(\delta^2-r), \tag{21}$$

and

$$R=\frac{\delta}{pqA}\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}\{p^2a^rb^{1-r'}-q^2b^ra^{1-r'}\}\times h(\delta,r,r')F(\delta^2-r). \tag{22}$$

By making use of Eqs. (17)—(22) Eqs. (15) and (16) may

 0022-2488/78/1906-1376$1.00

now be written in the form

$$\sigma\frac{\partial\omega}{\partial x}=-\kappa\left\{\frac{\partial Q}{\partial x}+\frac{1}{2}\left(\frac{1}{\sigma}+\sigma\right)\frac{\partial R}{\partial x}\right\}$$

and

$$\sigma\frac{\partial\omega}{\partial y}=-\kappa\left\{\frac{\partial Q}{\partial y}+\frac{1}{2}\left(\frac{1}{\sigma}+\sigma\right)\frac{\partial R}{\partial y}\right\},$$

which give

$$\sigma\omega=-\kappa[Q+(1/2)\{(1/\sigma)+\sigma\}R]+\text{const.} \tag{23}$$

The constant in Eq. (23) is determined to be $(\sigma/2)\{(1/\sigma)+\sigma\}(m/q)$ so as to get $\omega\to 0$ when $q\to 0$. Then we get Eqs. (3) and (7) with the help of Eq. (5) of the former of Ref. 13.

Next, the metric function γ satisfies

$$2\frac{\partial\gamma}{\partial a}=\frac{1}{A}\frac{\partial A}{\partial a}-\frac{\delta^2}{a-b}\quad\text{and}\quad 2\frac{\partial\gamma}{\partial b}=\frac{1}{A}\frac{\partial A}{\partial a}+\frac{\delta^2}{a-b},$$

also in the charged version, from which we get

$$2\gamma=\ln A-\delta^2\ln(a-b)+\text{const.} \tag{24}$$

The constant in Eq. (24) is determined to be $-\ln p^{2\delta}$ so as to get $\gamma\to 0$, when $x\to\infty$. Thus we obtain Eq. (4). From Eq. (8) of Ref. 4 we obtain Eqs. (10) and (11) with the help of Eqs. (17)–(20).

The present solutions may now be written in the Boyer–Lindquist form

$$ds^2=\frac{\Sigma^2}{\Delta}dr^2+\Sigma^2 d\theta^2+\frac{B\Delta\sin^2\theta}{A}d\phi^2-\frac{A}{B}\left(dt-\frac{2\alpha C\sin^2\theta}{\sigma^2 A}d\phi\right)^2, \tag{25}$$

where

$$r=(mp\sigma/\delta)x+m,\quad\cos\theta=y,\quad\rho=\Delta^{1/2}\sin\theta,$$

$$z=(r-m)\cos\theta,\quad\Delta=(r-m)^2-\{(m^2-\alpha^2-e^2)/\delta^2\}$$

$$=(mp\sigma/\delta)^2 a,$$

$$\Sigma^2=\frac{Bm^{2\delta}\sigma^{2\delta}}{\delta^2(m^2-\alpha^2-e^2)^{\delta-1}}\left(\frac{\delta^2\Delta}{m^2-\alpha^2-e^2}+\sin^2\theta\right)^{1-\delta^2},$$

and

$$(mp\sigma)^2=m^2-\alpha^2-e^2.$$

The area of the surface $x=1$ (event horizon for odd δ) is

$$S=(4\pi m^2/\delta)(q/p)^{\delta-1}(1+2p\sigma+\sigma^2)\sqrt{(-1)^{\delta-1}}.$$

3. BONNOR-MPST-WANG FAMILY OF SOLUTIONS

In the present section we shall present the Bonnor[2]–Misra–Pandey–Srivastava–Tripathi[5]–Wang[7] family of polarized charge solutions to Einstein–Maxwell equations for arbitrary integer distortion parameter δ. It should be stressed that the K–TS family of solutions and the B–MPST–W family of solutions are mathematically identical, though of course physically quite different. One interprets the parameter q as representing the angular momentum $J=m^2q$ for the K–TS family of solutions and the dipole moment $P=-2m^2q$ for the B–MPST–W family of solutions.

From the vacuum Einstein equations $R_{\mu\nu}=0$ with the line element given in Eq. (1) one obtains the following set of equations:

$$\Delta f-\frac{1}{f}\left(\frac{\partial f}{\partial z}\right)^2-\frac{1}{f}\left(\frac{\partial f}{\partial\rho}\right)^2+\frac{f^3}{\rho^2}\left(\frac{\partial\omega}{\partial z}\right)^2+\frac{f^3}{\rho^2}\left(\frac{\partial\omega}{\partial\rho}\right)^2=0, \tag{26}$$

$$\Delta\omega-\frac{2}{\rho}\frac{\partial\omega}{\partial\rho}+\frac{2}{f}\frac{\partial f}{\partial z}\frac{\partial\omega}{\partial z}+\frac{2}{f}\frac{\partial f}{\partial\rho}\frac{\partial\omega}{\partial\rho}=0, \tag{27}$$

$$\frac{\partial\gamma}{\partial z}-\frac{\rho}{2f^2}\frac{\partial f}{\partial z}\frac{\partial f}{\partial\rho}+\frac{f^2}{2\rho}\frac{\partial\omega}{\partial z}\frac{\partial\omega}{\partial\rho}=0, \tag{28}$$

$$\Delta\gamma+\frac{1}{2f^2}\left(\frac{\partial f}{\partial z}\right)^2+\frac{f^2}{2\rho^2}\left(\frac{\partial\omega}{\partial\rho}\right)^2=0, \tag{29}$$

and

$$\Delta\gamma-\frac{2}{\rho}\frac{\partial\gamma}{\partial\rho}+\frac{1}{2f^2}\left(\frac{\partial f}{\partial\rho}\right)^2+\frac{f^2}{2\rho^2}\left(\frac{\partial\omega}{\partial z}\right)^2=0, \tag{30}$$

where

$$\Delta=\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}.$$

The twist potential Ω is defined by[15]

$$\frac{\partial\omega}{\partial z}=\frac{\rho}{f^2}\frac{\partial\Omega}{\partial\rho}\quad\text{and}\quad\frac{\partial\omega}{\partial\rho}=-\frac{\rho}{f^2}\frac{\partial\Omega}{\partial z}. \tag{31}$$

Equations (26) and (31) are the Ernst equation.[15]

Now we discuss static, polarized-charged, axisymmetric, asymptotically flat, and exact solutions to the Einstein–Maxwell equations $G_{\mu\nu}=8\pi T_{\mu\nu}$ under conditions where space is empty of everything except a source-free electromagnetic field

$$G_{\mu\nu}=R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=8\pi T_{\mu\nu},$$

$$4\pi T_{\mu\nu}=F_{\mu\alpha}g^{\alpha\beta}F_{\nu\beta}-\tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}.$$

Assume that, this is a key point, the line element for the present problem may now be written in the form

$$ds^2=f^{-2}[e^{8\gamma}(dz^2+d\rho^2)+\rho^2 d\phi^2]-f^2dt^2, \tag{32}$$

where metric functions f and γ are given in Eqs. (2) and (4) under the constraint $\sigma=1$, respectively. Assume again that, this is another key point, the function ω given in Eq. (3) under $\sigma=1$, now not being the metric function, may now define the nonvanishing electromagnetic field tensors in the form

$$F_{tz}=-\frac{f^2}{\rho}\frac{\partial\omega}{\partial\rho}\sin\beta=\frac{\partial\Omega}{\partial z}\sin\beta, \tag{33}$$

$$F_{t\rho}=\frac{f^2}{\rho}\frac{\partial\omega}{\partial z}\sin\beta=\frac{\partial\Omega}{\partial\rho}\sin\beta, \tag{34}$$

$$F_{\rho\phi}=-\frac{\partial\omega}{\partial\rho}\cos\beta=\frac{\rho}{f^2}\frac{\partial\Omega}{\partial z}\cos\beta, \tag{35}$$

and

$$F_{\phi z}=\frac{\partial\omega}{\partial z}\cos\beta=\frac{\rho}{f^2}\frac{\partial\Omega}{\partial\rho}\cos\beta, \tag{36}$$

where β is an arbitrary angle associated with a duality rotation.[16]

It follows, as we defined them so as to do, that the Einstein curvature tensor $G_{\mu\nu}$ and the electromagnetic field tensor $F_{\mu\nu}$ given in Eqs. (33)–(36) satisfy the

Einstein—Maxwell equations with the help of Eqs. (26), (28), (29), (30), and (31). Therefore the B—MPST—W family of solutions with arbitrary distortion parameter δ is given by Eqs. (32)—(36) with Eqs. (2)—(7) under the constraint $\sigma = 1$.

The asymptotic form of Ω and ω are $\Omega \approx -2m^2 q \cos\theta / r^2$ and $\omega \approx -2m^2 q \sin^2\theta / r$, from which one interprets $P = -2m^2 q$ as the dipole moment.

ACKNOWLEDGMENT

The author would like to thank Professor S. Hori for many interesting discussions.

[1]E. T. Newman, E. Couch, K. Chinapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[2]W. B. Bonnor, Z. Physik **190**, 444 (1966).
[3]B. Carter, Phys. Rev. **174**, 1559 (1968).
[4]F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[5]R. M. Misra, D. B. Pandey, D. C. Srivastava, and S. N. Tripathi, Phys. Rev. D **7**, 1587 (1973).
[6]F. J. Ernst, Phys. Rev. D **7**, 2520 (1973).
[7]M. Y. Wang, Phys. Rev. D **9**, 1835 (1974).
[8]K. C. Das and S. Banerji, Phys. Lett. A **50**, 409 (1975).
[9]C. Reina and A. Treves, J. Math. Phys. **16**, 834 (1975).
[10]G. Önengüt and M. Serdaroğlu, Nuovo Cimento B **27**, 213 (1975).
[11]R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[12]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972); and Prog. Theor. Phys. **50**, 95 (1973).
[13]M. Yamazaki, J. Math. Phys. **18**, 2502 (1977); and Prog. Theor. Phys. **57**, 1951 (1977).
[14]S. Hori (to be published).
[15]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[16]C. W. Misner and J. A. Wheeler, Ann. Phys. (N. Y.) **2**, 525 (1957).

Simple analytic expressions for the Coulomb off-shell Jost functions

H. van Haeringen

Natuurkundig Laboratorium der Vrije Universiteit, Amsterdam, The Netherlands
(Received 15 December 1977)

The off-shell Jost functions have been introduced by Fuda and Whiting. We give simple closed expressions for $f_{C,l}(k,q)$, the off-shell Jost functions for the Coulomb potential, and we derive their connection with the ordinary Coulomb Jost functions $f_{C,l}(k)$.

The concept of a Jost[1] function is well known in the theory of nonrelativistic two-body scattering.[2] Fuda and Whiting[3] have introduced an off-shell generalization of the Jost function $f_l(k)$, which they call the off-shell Jost function $f_l(k,q)$. This function is closely related to the half-off-shell T matrix $T_l(k,q;k^2)$ which plays such an important role in few-body calculations.

For a short-range potential the on-shell limit for $q\to k$ of $f_l(k,q)$ exists and is equal to the ordinary Jost function. This is not true for long-range potentials such as the Coulomb potential. Recently we have derived the $l=0$ Coulomb off-shell Jost function and studied its connection with the Hulthén off-shell Jost function.[4]

In this paper we give an exact closed expression for the Coulomb off-shell Jost function $f_{C,l}(k,q)$ for all l. It follows from this expression that $f_{C,l}(k,q)$ is singular at $q=k$. Below we shall prove that [see Eq. (40.7) of Ref. 4]

$$\lim_{q\to k}\omega f_{C,l}(k,q)=f_{C,l}(k),\quad l=0,1,2,\ldots, \tag{1}$$

with

$$\omega\equiv\left(\frac{q-k}{q+k}\right)^{i\gamma}\frac{\exp(\pi\gamma/2)}{\Gamma(1+i\gamma)}, \tag{2}$$

where γ is Sommerfeld's parameter.

We expect to be able to prove in the near future that a similar relation holds for the off-shell Jost function $f_{C+S,l}(k,q)$ for a potential which is the sum of the Coulomb potential V_C and an arbitrary short-range potential V_S, namely

$$\lim_{q\to k}\omega f_{C+S,l}(k,q)=f_{C+S,l}(k), \tag{3}$$

with the *same* ω of Eq. (2).

Our starting point is an integral representation obtained by Fuda [Ref. 5, Eq. (25)] which we rewrite as follows,

$$f_{C,l}(k,q)=1+\tfrac{1}{2}\pi q(q/k)^l\,{}_0\langle ql\downarrow|V_{C,l}|kl+\rangle_C f_{C,l}(k). \tag{4}$$

Here $|kl+\rangle_C$ is the (outgoing) Coulomb scattering state and

$$f_{C,l}(k)=\exp(\pi\gamma/2)\Gamma(l+1)/\Gamma(l+1+i\gamma) \tag{5}$$

is the Coulomb Jost function. Furthermore (see Ref. 6),

$${}_0\langle ql\downarrow|r\rangle=(-)^l\langle r|ql\downarrow\rangle_0=(2/\pi)^{1/2}i^{-l}h_l^{(+)}(qr).$$

By applying a number of relations existing between various special functions we have been able to derive the following exact expression,

$$\begin{aligned}{}_0\langle ql\downarrow|V_{C,l}|kl+\rangle_C&={}_C\langle kl-|V_{C,l}|ql\downarrow\rangle_0\\&=[2i\gamma/(\pi q)]f_{C,l}^{-1}(k)\Gamma(l+1)\\&\times\sum_{m=0}^{l}[\Gamma(l+m+1)/\Gamma(m+1)](k/q)^m z^{-l-m}\\&\times\Bigg[\frac{\Gamma(m-i\gamma)}{\Gamma(l+1+m)\Gamma(l+1-i\gamma)}\\&\times{}_2F_1(-m-l,\ i\gamma-l;\ 1+i\gamma-m;\ 1-z)\\&+(1-z)^{m-i\gamma}\frac{\Gamma(i\gamma-m)}{\Gamma(l+1-m)\Gamma(l+1+i\gamma)}\\&\times{}_2F_1(m-l,\ -i\gamma-l;\ 1-i\gamma+m;1-z)\Bigg],\end{aligned} \tag{6}$$

with $z=2k/(q+k)$. Note that the second term between the square brackets is, apart from the factor $(1-z)^{m-i\gamma}$, just equal to the first term if one replaces m by $-m$ and γ by $-\gamma$.

We point out that both hypergeometric series ${}_2F_1$ occurring in Eq. (6) are terminating ones. Although Eq. (6) looks somewhat complicated, it has the important property that the branch-cut singularity, which is contained in the factor $(1-z)^{-i\gamma}$, can be split off. This has the advantage of revealing the analytic structure of the quantity ${}_0\langle ql\downarrow|V_{C,l}|kl+\rangle_C$.

With the help of some further manipulations we have reduced Eq. (6) to the following equivalent, more convenient form,

$$\begin{aligned}{}_0\langle ql\downarrow|V_{C,l}|kl+\rangle_C&=\frac{2}{\pi q}c_{l\gamma}f_{C,l}^{-1}(k)x^{-l}\\&\times\left[-A_l(x^2;\gamma^2)+x^l\left(\frac{q+k}{q-k}\right)^{i\gamma}P_l^{(-i\gamma,i\gamma)}(u)\right],\end{aligned} \tag{7}$$

where $x=q/k$, $u=(q^2+k^2)/(2qk)$ and we have used the abbreviation,

$$\begin{aligned}c_{l\gamma}&\equiv\binom{l+i\gamma}{l}^{-1}\binom{l-i\gamma}{l}^{-1}\\&=\frac{\Gamma^2(l+1)\Gamma(1+i\gamma)\Gamma(1-i\gamma)}{\Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)}=\prod_{n=1}^{l}(1+\gamma^2/n^2)^{-1}.\end{aligned} \tag{8}$$

Furthermore, $P_l^{(-i\gamma,i\gamma)}$ is Jacobi's polynomial and A_l is a certain *polynomial of two variables*. Its degree is, in both variables separately, equal to l and it has real coefficients. For $l=0$, 1, and 2 we have obtained,

$$A_0(x^2;\gamma^2)=1,$$

$$A_1(x^2;\gamma^2)=\tfrac{1}{2}(x^2+1+2\gamma^2),$$

$$A_2(x^2;\gamma^2)=\tfrac{1}{8}[3x^4+2x^2(1+\gamma^2)+3+2\gamma^2(4+\gamma^2)].$$

 0022-2488/78/1906-1379$1.00

For general values of l we have proved the following important properties,

$$A_l(1;\gamma^2)=c_{l\gamma}^{-1}, \tag{9}$$

which gives the on-shell value, and for the case of vanishing potential strength,

$$A_l(x^2;0)=x^l P_l(u),$$

where P_l is Legendre's polynomial. In view of Eqs. (4) and (7) the Coulomb off-shell Jost function is given by

$$f_{C,l}(k,q)=1+c_{l\gamma}\left[-A_l(x^2;\gamma^2)+x^l\left(\frac{q+k}{q-k}\right)^{i\gamma}P_l^{(-i\gamma,i\gamma)}(u)\right]. \tag{10}$$

The proof of Eq. (1) now follows from Eqs. (9) and (10) and the equalities

$$P_l^{(-i\gamma,i\gamma)}(1)=\binom{l-i\gamma}{l},$$

$$f_{C,l}(k)=\binom{l-i\gamma}{l}c_{l\gamma}f_{C,0}(k).$$

ACKNOWLEDGMENT

This investigation forms a part of the research program of the Foundation for Fundamental Research of Matter (FOM), which is financially supported by the Netherlands Organization for Pure Scientific Research (ZWO).

[1]R. Jost, Helv. Phys. Acta 20, 256 (1947).
[2]R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[3]M.G. Fuda and J.S. Whiting, Phys. Rev. C 8, 1255 (1973).
[4]H. van Haeringen, "Fuda's off-shell Jost function for Coulomb, Hulthén and Eckart potentials and limiting relations," preprint (1977).
[5]M.G. Fuda, Phys. Rev. C 14, 37 (1976).
[6]H. van Haeringen, Nuovo Cimento B 34, 53 (1976).

Topological cohesion

K. Cahill and N. Stoltzfus

Departments of Mathematics and of Physics & Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803
(Received 19 December 1977)

It is shown that suitably regular, finite-energy solutions of the Yang–Mills–Higgs equations are nondissipative whenever their initial data are topologically significant.

It is well known that under certain conditions topologically stable field configurations can arise as regular solutions of the classical field equations that describe the locally gauge-invariant interaction of scalar Higgs mesons with vector Yang—Mills mesons.[1] The purpose of the present paper is to show that under very general conditions every suitably regular, finite-energy solution of the field equations for such a system is nondissipative provided that its initial data are topologically significant. This effect, which we have called topological cohesion, means that topological solitons commonly occur in such theories. It also suggests that they may exhibit some of the resiliency associated with the stricter meaning of the term soliton.[2]

The interaction of the scalar mesons ϕ^i with the vector mesons A^a_μ will be assumed to be described by a Lagrange density L that is invariant, in the usual Yang–Mills way, under the local action of a compact Lie group G. We shall not need the explicit form of L in what follows. It will be sufficient to know that the potential or self-interaction $V(\phi)$ of the Higgs mesons is a continuous, nonnegative, G-invariant function of the ϕ^i and that the points ϕ on which $V(\phi)$ assumes its minimum value, zero, form a smooth compact manifold M. It will also be assumed that the Hamilton density H derived from L is bounded below by V, as is usually the case, and that $V(\phi)$ itself is bounded below by some fixed positive number for sufficiently large $\|\phi\|$.

Since the vacuum manifold M is smooth, it follows[3] that there exists a (G-invariant) tubular neighborhood $N(M) \supset M$ and a continuous map r from $N(M)$ into M that is the identity on M itself, i.e., for ϕ in M, $r(\phi)=\phi$. Because $N(M)$ is open, its complement is closed; and so on it the continuous function $V(\phi)$, which avoids zero for large $\|\phi\|$, is bounded below by some fixed positive number ϵ. Suppose now that $\phi^i(x,t)$ and $A^a_\mu(x,t)$ form a solution of the field equations with finite energy $E = \int d^n x H(x,t)$. Then since $H \geq V \geq \epsilon$ on the complement of $N(M)$, the fields $\phi^i(x,t)$ must for all times t lie inside $N(M)$ for all points x that lie outside a region $P(t)$ whose volume $v(P(t))$ is never larger than E/ϵ. If the region $P(t)$ does not have unbounded horns or whiskers, then it can be put inside a sphere of radius $R(t)$ centered about the origin. In this paper we use the term *suitably regular* to denote a finite-energy solution $\phi^i(x,t)$ and $A^a_\mu(x,t)$ that is continuous as a function of x and t and for which the radius $R(t)$ is bounded on compact time intervals. It is likely[4] that reasonable initial data, $\phi^i(x,0)$, $A^a_\mu(x,0)$ and $\dot{\phi}^i(x,0)$, $\dot{A}^a_\mu(x,0)$, lead to solutions that are suitably regular for positive t, at least for potentials $V(\phi)$ that are physically admissible.

If now $\phi^i(x,t)$ is suitably regular, then the point $\phi(\rho\hat{x},t)$, where $\hat{x}=x/\|x\|$, lies in the neighborhood $N(M)$ for all $\rho \geq R(t)$ and $t \geq 0$. The function $r(\phi(\rho\hat{x},t))$ is then, for such fixed t and ρ, a continuous map from the sphere S^{n-1} into M and therefore falls into one of the homotopy classes of $[S^{n-1},M]$. The function $r(\phi(\rho\hat{x},t))$, for fixed t, is also a continuous map from $S^{n-1}\times[R(t),\infty)$ into M. Thus it is a homotopy between the maps $r(\phi(\rho\hat{x},t))$ and $r(\phi(\rho'\hat{x},t))$ for all pairs ρ, $\rho' \geq R(t)$. The homotopy class of ϕ therefore is independent of ρ. It is also independent of t, since for fixed ρ, greater than the upper bound of $R(t)$ on the compact interval $[0,t]$, the function $r(\phi(\rho\hat{x},t))$ is a continuous map from $S^{n-1}\times[0,t]$ into M, which makes it a homotopy between $r(\phi(\rho\hat{x},0))$ and $r(\phi(\rho\hat{x},t))$. In this paper a suitably regular, finite-energy solution is said to possess *topologically significant* initial data if this homotopy class $[\phi]$ is nontrivial. For such solutions, the name of the class $[\phi]$ may be viewed as a conserved topological charge.

Suppose now that $\phi(x,t)$, $A_\mu(x,t)$ is a suitably regular, finite-energy solution with topologically significant initial data. Then for all $t \geq 0$, there is some point $x(t)$ for which $\phi(x(t),t)$ lies outside the tubular neighborhood $N(M)$. For if at some time $t \geq 0$, $\phi(x(t),t)$ were in $N(M)$ for all x in R^n, then the function $r(\phi(\rho\hat{x},t))$ for that value of t would be a continuous map from $S^{n-1}\times[0,\infty)$ into M. It would then be a homotopy between $r(\phi(0,t))$ which is in the trivial class and $r(\phi(\rho\hat{x},t))$ which or $\rho \geq R(t)$ is in $[\phi]$ assumed nontrivial. This contradiction means that for all $t \geq 0$ there is some point $x(t)$ for which $\phi(x(t),t)$ lies outside $N(M)$. But at that point $x(t)$ the energy density $H(x(t),t) \geq V(\phi(x(t),t)) \geq \epsilon$ since V is bounded below by ϵ on the complement of $N(M)$. Every suitably regular, finite-energy solution with topologically significant initial data is therefore nondissipative.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under contract MCS 77-02621 for which we are grateful.

[1] S. Coleman, in *New Phenomena in Subnuclear Physics* (Erice 1975, Part A, ed. Zichichi) (Plenum, New York, 1977).
[2] A. Scott, F. Chu, and D. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[3] G. Bredon, *Introduction to Compact Transformation Groups* (Academic, New York, 1972), p. 85, Lemma 5.1, Theorem 5.4.
[4] C. Parenti, F. Strocchi, and G. Velo, Nuovo Cimento B **39**, 147 (1977).

 0022-2488/78/1906-1381$1.00

Simple proof of no position operator for quanta with zero mass and nonzero helicity

Thomas F. Jordan

Department of Physics, University of Minnesota, Duluth, Duluth, Minnesota 55812 [a]
and Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627 [b]
(Received 18 May 1977)

That there is no Newton–Wigner position operator in an irreducible unitary representation of the Poincaré group for zero mass and either variable helicity or fixed nonzero helicity is proved formally using familiar operator algebra and transformations and bringing out relevant properties of operators in the representation. In the case of fixed helicity, a pair of irreducible representations with opposite helicities is considered, so the parity operator is defined; then the correct parity transformation is assumed for the position operator.

1. INTRODUCTION

Although it is well known that there is no Newton–Wigner position operator in an irreducible unitary representation of the Poincaré group for zero mass and nonzero helicity,[1–3] the rigorous proof uses elegant mathematical methods not familiar to most physicists.[2] Here we give a formal proof using operator algebra and transformations familiar in ordinary quantum mechanics and bringing out relevant properties of operators in the representations.

We consider irreducible unitary representations of the Poincaré group for zero mass and both fixed helicity, that is "discrete spin," and variable helicity, that is "continuous spin."[4] In the case of fixed helicity we consider a pair of irreducible representations with opposite helicities, as is usual for photons, so the parity operator is defined. Then we require the correct parity transformation for the position operator.

In all cases we require the position operator to transform correctly for space translations, rotations, and time reversal. We assume the position operator is Hermitian, and we assume its components commute with each other.

2. GENERATORS

Let $\mathbf{P}$ denote the generator for space translations. We consider representations for zero mass and positive energy so the Hamiltonian, the generator for time translations, is

$$H=(\mathbf{P}^2)^{1/2}. \tag{2.1}$$

For the rotation and Lorentz-transformation generators $\mathbf{J}$ and $\mathbf{K}$ we use the form written by Lomont and Moses.[5] They take the little group to be the subgroup of the Lorentz group that leaves unit momentum in the x direction unchanged. For the representation of the little group, the two-dimensional Euclidean group, they have generators T_2 and T_3 which commute with each other, and S, the helicity, whose commutation relations with T_2 and T_3 are those of a generator for rotations around the x axis with the y and z components of a vector. The irreducible unitary representation of the Poincaré group is spanned by eigenkets $|\mathbf{p},s\rangle$ of $\mathbf{P}$ and S. In a representation with variable helicity, the eigenvalues of S are either all the integers $s=0,\pm1,\pm2,\cdots$ for a "single-valued" representation or all the half-integers $s=\pm\frac{1}{2},\pm\frac{3}{2},\cdots$ for a "double-valued" representation, and

$$(T_3\pm iT_2)|\mathbf{p},s\rangle=r|\mathbf{p},s\mp1\rangle, \tag{2.2}$$

where r is a positive number characteristic of the representation. In a representation with fixed helicity, S has just a single integer or half-integer value, and T_2, T_3 are zero. Let

$$\mathbf{M}=(1,P_2/(H+P_1),P_3/(H+P_1))$$
$$=(H\hat{x}+\mathbf{P})/(H+\mathbf{P}\cdot\hat{x}), \tag{2.3}$$

$$\mathbf{N}=(0,P_3/(H+P_1),-P_2/(H+P_1))$$
$$=\mathbf{P}\times\hat{x}/(H+\mathbf{P}\cdot\hat{x}), \tag{2.4}$$

$$\mathbf{E}_2=(P_3/H,P_2P_3/H(H+P_1),-P_2^2/H(H+P_1)-P_1/H), \tag{2.5}$$

$$\mathbf{E}_3=(-P_2/H,P_3^2/H(H+P_1)+P_1/H,-P_2P_3/H(H+P_1)). \tag{2.6}$$

The rotation and Lorentz-transformation generators are[5]

$$\mathbf{J}=\mathbf{Q}\times\mathbf{P}+\mathbf{M}S, \tag{2.7}$$

$$\mathbf{K}=\tfrac{1}{2}(H\mathbf{Q}+\mathbf{Q}H)+\mathbf{N}S-H^{-1}(\mathbf{E}_3T_2-\mathbf{E}_2T_3). \tag{2.8}$$

Here $\mathbf{Q}$ is $i\nabla$ on momentum-space wavefunctions $\phi(\mathbf{p})$ with the ordinary inner product defined with the noninvariant $\int d^3p$.

The vectors $\mathbf{E}_2$, $\mathbf{E}_3$, and $\hat{P}=\mathbf{P}/H$ are orthonormal and $\hat{P}\times\mathbf{E}_2$ is $\mathbf{E}_3$, etc. The Pauli–Lubanski 4-vector is

$$W_0=\mathbf{P}\cdot\mathbf{J}=HS, \tag{2.9}$$

$$\mathbf{W}=H\mathbf{J}+\mathbf{P}\times\mathbf{K}=\mathbf{P}S+\mathbf{E}_2T_2+\mathbf{E}_3T_3. \tag{2.10}$$

We also use time reversal. To fit in with the other transformations, time reversal must be represented by an antiunitary antilinear operator that commutes with H and $\mathbf{K}$ and anticommutes with $\mathbf{P}$ and $\mathbf{J}$. Therefore, it commutes with W_0 and S and anticommutes with $\mathbf{W}$.

3. FIXED HELICITY

First we look for a position operator in an irreducible representation for zero mass and fixed helicity. For

[a] Permanent.
[b] 1976–77.

 0022-2488/78/1906-1382$1.00

zero helicity the Newton–Wigner position operator is **Q**. It transforms as a position operator should for translations, rotations, and time reversal. It is Hermitian, and its components commute with each other. We can easily see there is no other position operator. Suppose **R** is a position operator. Let

$$\mathbf{R}=\mathbf{Q}+\mathbf{F}.$$

Then **F** must be Hermitian, invariant for translations, a vector for rotations, and invariant for time reversal. Since **F** commutes with **P**, and **P** is a complete set of commuting operators, **F** must be a function of **P**. Since **F** is a vector for rotations, it must be the form

$$\mathbf{F}=f(\mathbf{P}^2)\mathbf{P}$$

with some real function f of $\mathbf{P}^2$. Then time reversal changes **F** to $-\mathbf{F}$ so, since it is invariant for time reversal, **F** must be zero.

Next we consider an irreducible representation for fixed nonzero helicity. The operator

$$H^{-1}(\mathbf{K}-i\mathbf{P}/2H)=\mathbf{Q}+H^{-1}\mathbf{N}S$$

is Hermitian and transforms as a position operator should for translations, rotations, and time reversal (because $H^{-1}\mathbf{N}S$ is invariant for translations, **K** and $i\mathbf{P}$ are vectors for rotations and invariant for time reversal, and H is invariant for rotations and time reversal). Suppose **R** is a position operator. Let

$$\mathbf{R}=\mathbf{Q}+H^{-1}\mathbf{N}S+\mathbf{F}.$$

Then **F** must be Hermitian, invariant for translations, a vector for rotations, and invariant for time reversal. As above, this implies that **F** is zero. Then the components of R do not commute. In fact

$$[R_j,R_k]=-i\epsilon_{jkl}P_lH^{-3}S.$$

Thus there is no position operator in an irreducible unitary representation of the Poincaré group for zero mass and fixed nonzero helicity.

Now we consider two irreducible representations with opposite helicities. We look for a position operator on the space spanned by eigenkets $|\mathbf{p},\lambda\rangle$ of P and helicity for two helicity values $\lambda=\pm|s|$. In place of S we have the helicity operator $|s|\Sigma_3$, where

$$\Sigma_3|\mathbf{p},\lambda\rangle=(\lambda/|s|)|\mathbf{p},\lambda\rangle. \tag{3.1}$$

We also use operators Σ_1 and Σ_2 defined by

$$\Sigma_1|\mathbf{p},\lambda\rangle=|\mathbf{p},-\lambda\rangle, \tag{3.2}$$

$$\Sigma_2=-i\Sigma_3\Sigma_1. \tag{3.3}$$

Parity and time-reversal operators that fit in with the other transformations are of the form[6]

$$P|\mathbf{p},\lambda\rangle=\omega(\lambda)^*\exp[-i\phi(\mathbf{p})\lambda]|-\mathbf{p},-\lambda\rangle,$$

$$T|\mathbf{p},\lambda\rangle=\eta(\lambda)^*\exp[i\phi(\mathbf{p})\lambda]|-\mathbf{p},\lambda\rangle,$$

where $\omega(\lambda)$ and $\eta(\lambda)$ are phase factors depending on λ and $\phi(\mathbf{p})$ is an angle depending on p with

$$\phi(-\mathbf{p})=\phi(\mathbf{p}).$$

Then

$$PT|\mathbf{p},\lambda\rangle=\omega(\lambda)^*\eta(\lambda)^*|\mathbf{p},-\lambda\rangle,$$

$$TP|\mathbf{p},\lambda\rangle=\omega(\lambda)\eta(-\lambda)^*|\mathbf{p},-\lambda\rangle.$$

These differ by a factor

$$\omega(\lambda)\omega(\lambda)\eta(\lambda)\eta(-\lambda)^*$$

This factor must be the same for $\lambda=\pm|s|$ because PT and TP, which represent the same transformation of space–time coordinates, can differ only by a phase factor. Therefore

$$\omega(\lambda)\omega(\lambda)\eta(\lambda)\eta(-\lambda)^*=\omega(-\lambda)\omega(-\lambda)\eta(-\lambda)\eta(\lambda)^*,$$

$$\omega(\lambda)\omega(\lambda)\omega(-\lambda)^*\omega(-\lambda)^*\eta(\lambda)\eta(\lambda)\eta(-\lambda)^*\eta(-\lambda)^*=1,$$

$$\omega(\lambda)\omega(-\lambda)^*\eta(\lambda)\eta(-\lambda)^*=\pm1.$$

The last is the factor by which

$$\Sigma_1PT|\mathbf{p},\lambda\rangle=\omega(\lambda)^*\eta(\lambda)^*|\mathbf{p},\lambda\rangle$$

differs from

$$PT\Sigma_1|\mathbf{p},\lambda\rangle=\omega(-\lambda)^*\eta(-\lambda)^*|\mathbf{p},\lambda\rangle,$$

and this factor is the same for $\lambda=\pm|s|$ so

$$\Sigma_1PT=\pm PT\Sigma_1.$$

Since PT anticommutes with i and with

$$\Sigma_3=\mathbf{P}\cdot\mathbf{J}/H|s|,$$

it follows from the definition (3.3) of Σ_2 that PT either commutes with both Σ_1 and Σ_2 or anticommutes with both Σ_1 and Σ_2.

The operator

$$H^{-1}(\mathbf{K}-i\mathbf{P}/2H)=\mathbf{Q}+H^{-1}\mathbf{N}|s|\Sigma_3$$

is Hermitian and transforms as a position operator should for translations, rotations, time reversal, and parity (because parity changes both P and K but not H). Suppose R is a position operator and let

$$\mathbf{R}=\mathbf{Q}+H^{-1}\mathbf{N}|s|\Sigma_3+\mathbf{D}, \tag{3.4}$$

Then D must be Hermitian, invariant for translations, a vector for rotations, and invariant for time reversal. Now we assume the position operator also transforms correctly for parity, which implies that D anticommutes with P. Then D anticommutes with PT.

Since D is Hermitian and translation-invariant, it must be of the form

$$\mathbf{D}=\mathbf{A}+\mathbf{B}\Sigma_1+\mathbf{C}\Sigma_2+\mathbf{F}\Sigma_3 \tag{3.5}$$

with A, B, C, F real functions of P.

Since D is also a vector for rotations, it must satisfy the commutation relation

$$(1/i)[\mathbf{D},\mathbf{P}\cdot\mathbf{J}]=\mathbf{P}\times\mathbf{D}.$$

Substituting $H|s|\Sigma_3$ for $\mathbf{P}\cdot\mathbf{J}$ and the form (3.5) for D, we get

$$2H|s|\mathbf{C}=\mathbf{P}\times\mathbf{B},\quad -2H|s|\mathbf{B}=\mathbf{P}\times\mathbf{C} \tag{3.6}$$

from the coefficients of Σ_1 and Σ_2. This implies

$$-(2|s|)^2\mathbf{B}=H^{-2}\mathbf{P}\times(\mathbf{P}\times\mathbf{B})=-\mathbf{B}$$

so we see B and C are zero if $|s|$ is not $\frac{1}{2}$.

In any case, since PT anticommutes with D and Σ_3, and either commutes with both Σ_1 and Σ_2 or anticommutes with both Σ_1 and Σ_2, it either anticommutes with both B and C or commutes with both B and C, which means B and C are either both odd or both even as functions of P. Then Eqs. (3.6) imply B and C are zero.

With its Σ_1, Σ_2 terms eliminated, R is reduced to a separate operator on each of the two irreducible representations. We have already showed there are no such operators. Thus there is no position operator for two irreducible unitary representations of the Poincaré group with zero mass and opposite fixed helicities.

4. VARIABLE HELICITY

Finally we look for a position operator R in an irreducible representation for zero mass and variable helicity. The operator

$$H^{-1}(\mathrm{K} - i\mathrm{P}/2H) = \mathrm{Q} + H^{-1}\mathbf{N}\mathrm{S} - H^{-2}(\mathrm{E}_3 T_2 - \mathbf{E}_2 T_3)$$

is Hermitian and transforms as a position operator should for translations, rotations, and time reversal (as before). Let

$$\mathrm{R} = \mathrm{Q} + H^{-1}\mathrm{NS} - H^{-2}(\mathrm{E}_3 T_2 - \mathrm{E}_2 T_3) + \mathrm{D}. \tag{4.1}$$

Then D must be Hermitian, invariant for translations, a vector for rotations, and invariant for time reversal.

Let

$$\mathrm{D} = A\hat{P} + B_2 \mathrm{E}_2 + B_3 \mathbf{E}_3, \tag{4.2}$$

From this and

$$\hat{P} \times \mathbf{W} = \mathbf{E}_3 T_2 - \mathbf{E}_2 T_3, \tag{4.3}$$

$$-\hat{P} \times (\hat{P} \times \mathrm{W}) = \mathbf{E}_2 T_2 + \mathbf{E}_3 T_3 \tag{4.4}$$

it follows that

$$\mathrm{D} \cdot \hat{P} = A, \quad \mathrm{D} \cdot \hat{P} \times \mathrm{W} = -B_2 T_3 + B_3 T_2,$$
$$-\mathrm{D} \cdot \hat{P} \times (\hat{P} \times \mathrm{W}) = B_2 T_2 + B_3 T_3.$$

These are translation-invariant rotational scalars. They commute with P and J. From Eq. (2.9) we see they commute with S. Then they are functions of P and S, because P and S are a complete set of commuting operators. Thus we have

$$(B_3 \pm iB_2)(T_3 \mp iT_2)|\mathbf{p}, s\rangle$$
$$= [B_2 T_2 + B_3 T_3 \mp i(-B_2 T_3 + B_3 T_2)]|\mathbf{p}, s\rangle$$
$$= f_\pm(\mathbf{p}, s)|\mathbf{p}, s\rangle$$

with some functions $f_\pm$. From Eqs. (2.2) it follows that

$$B_3 \pm iB_2 = (1/r^2) f_\pm(\mathbf{P}, S)(T_3 \pm iT_2)$$

or

$$B_2 = (1/2r^2)(f_+ + f_-)T_2 - (i/2r^2)(f_+ - f_-)T_3,$$

$$B_3 = (i/2r^2)(f_+ - f_-)T_2 + (1/2r^2)(f_+ + f_-)T_3$$

so

$$\mathrm{D} = A\hat{P} + iB(\mathbf{E}_2 T_2 + \mathbf{E}_3 T_3) + C(\mathbf{E}_3 T_2 - \mathrm{E}_2 T_3), \tag{4.5}$$

with A, B, C functions of P and S.

For D to be a vector for rotations, A, B, and C must be rotational scalars, because they multiply rotational vectors that are orthogonal, as we can see from Eqs. (4.3) and (4.4). From Eq. (2.9) we see S is a rotational scalar. Therefore, A, B, C must be functions of $\mathbf{P}^2$ and S.

For D to be unchanged by time reversal, A must be imaginary and B, C real, because $\hat{P}$ is changed, P^2 is not, and S is not, as we can see from Eq. (2.9), and iB multiplies a vector that is changed and C a vector that is not, as we can see from Eqs. (4.3), (4.4), and (2.10).

For D to be Hermitian, A must be zero and B, C must satisfy

$$-iT_2 B - T_3 C = iBT_2 - CT_3,$$
$$-iT_3 B + T_2 C = iBT_3 + CT_2,$$

which are equivalent to the commutator–anticommutator relations

$$\{B, T_3 \pm iT_2\} = \pm[C, T_3 \pm iT_2].$$

From Eqs. (2.2) we can see this implies

$$B(s-1) + B(s) = C(s-1) - C(s),$$
$$B(s+1) + B(s) = C(s) - C(s+1) \tag{4.6}$$

for the functions of the eigenvalues s that define B and C as functions of S. It follows that

$$B(s-1) - B(s+1) = C(s-1) + C(s+1) - 2C(s),$$
$$C(s-1) - C(s+1) = B(s-1) + B(s+1) + 2B(s). \tag{4.7}$$

The dependence of B and C on P is suppressed here because it plays no role.

From Eqs. (2.2) we see that

$$T_2^2 + T_3^2 = r^2. \tag{4.8}$$

From Eqs. (2.2) also it follows[7] that for a function f of P and S

$$[f(S), T_2] = \tfrac{1}{2}[2f(S) - f(S+1) - f(S-1)]T_2$$
$$+ (i/2)[f(S+1) - f(S-1)]T_3,$$
$$[f(S), T_3] = \tfrac{1}{2}[2f(S) - f(S+1) - f(S-1)]T_3$$
$$- (i/2)[f(S+1) - f(S-1)]T_2. \tag{4.9}$$

Again the P dependence is suppressed because it plays no role. We have

$$\mathrm{R} = \mathrm{Q} + H^{-1}\mathbf{N}\mathrm{S} + iB(\mathbf{E}_2 T_2 + \mathrm{E}_3 T_3) + (C - H^{-2})(\mathbf{E}_3 T_2 - \mathbf{E}_2 T_3).$$

Let

$$[R_j, R_k] = i\epsilon_{jkl} G_l.$$

Using Eqs. (4.7)–(4.9), we find

$$\mathrm{G} = i2B' H(\mathbf{E}_3 T_2 - \mathbf{E}_2 T_3) - 2(C - H^{-2})' H(\mathbf{E}_2 T_2 + \mathbf{E}_3 T_3)$$
$$+ iBH^{-1}(\mathbf{E}_3 T_2 - \mathbf{E}_2 T_3) - (C - H^{-2})H^{-1}(\mathrm{E}_2 T_2 + \mathbf{E}_3 T_3)$$
$$+ \mathrm{P}H^{-1}[-H^{-2}\mathrm{S} + 2r^2 B(C - H^{-2})], \tag{4.10}$$

where $'$ denotes a derivative with respect to $\mathbf{P}^2$. Since we assume the different components of R commute with each other, G must be zero. The terms involving

$\mathbf{E}_2T_2 + \mathbf{E}_3T_3$ and $\mathbf{E}_3T_2 - \mathbf{E}_2T_3$ must be separately zero because, as we already noted, these vectors are orthogonal to each other and to P. This determines the dependence of B and $C - H^{-2}$ on $\mathbf{P}^2$; they must be proportional to H^{-1}. Let

$$B = r^{-1}H^{-1}b(S), \quad C - H^{-2} = r^{-1}H^{-1}c(S).$$

For G to be zero we must have

$$2b(S)c(S) = S \tag{4.11}$$

or

$$2b(s)c(s) = s \tag{4.12}$$

for the functions of the eigenvalues s that define b and c as functions of S. Combining Eqs. (4.6) and (4.12), and solving the resulting quadratic equation, we find

$$2c(s+1) = c(s) - b(s) \pm [c(s)^2 + b(s)^2 - 3s - 2]^{1/2}.$$

For real $c(s+1)$ we must have

$$c(s^2) + b(s)^2 \geq 3s + 2. \tag{4.13}$$

But from Eqs. (4.6) and (4.12) we have also

$$c(s+1) + b(s+1) = c(s) - b(s),$$

$$c(s+1)^2 + b(s+1)^2 + s + 1 = c(s)^2 + b(s)^2 - s,$$

which shows that $c(s)^2 + b(s)^2$ decreases for increasing positive s so that sooner or later Eq. (4.13) is not satisfied. Thus there is no position operator for an irreducible unitary representation of the Poincaré group with zero mass and variable helicity.

[1] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[2] A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).
[3] J. M. Jauch and C. Piron, Helv. Phys. Acta **40**, 559 (1967); W. O. Amrein, Helv. Phys. Acta **42**, 149 (1969). A Hermitian position operator with commuting components has been constructed for the photon on a null plane, "in its own reference system," by J. Bertrand, Nuovo Cimento A **15**, 281 (1973). See also E. Angelopoulos, F. Bayen and M. Flato, Phys. Scr. **9**, 173 (1974).
[4] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[5] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962).
[6] S. Weinberg, Phys. Rev. **134**, B882 (1964).
[7] T. F. Jordan, **19**, 247 (1978).

Solution of a modified Lamé equation with an integral term[a)]

P. L. Hagelstein[b)]

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and Lawrence Livermore Laboratory, Livermore, California 94550
(Received 5 December 1977)

We consider an equation which occurs in the stability analysis of a passively modelocked laser system in which the pulses overlap. The equation is related to a Lamé equation and can be written

$$su(x)=\{d^2/dx^2-[(2-m)-6\,\mathrm{dn}^2(x,m)]\}u(x)$$
$$-\epsilon\int_{-K(m)}^{K(m)}u(x')\,\mathrm{dn}(x',m)dx'(1+\nu^2d^2/dx^2)\,\mathrm{dn}(x,m),$$

where s is an eigenvalue and ϵ and ν are arbitrary parameters which need not be small. The eigenfunctions satisfy periodic boundary conditions $u(x+2K(m))=u(x)$. The eigenvalue s is proportional to the rate of growth of perturbations of the steady state pulse, and so we focus on finding the conditions under which no eigenvalue s is positive. We find that

$$\epsilon=2/\{\xi(m)-\nu^2[2E(m)-(2-m)\xi(m)]\}$$

marks the stability boundary, where $E(m)$ is the elliptic integral of the second kind and $\xi(m)$ is a function defined within the paper. This result is exact.

I. INTRODUCTION AND BACKGROUND

Within the framework of the fast absorber model of passive laser modelocking[1] one can find analytical solutions in the case of overlapping pulses[2,3] which can be expressed in terms of Jacobian elliptic functions. The dnoidal solutions lend themselves to a description of the modelocked waveform from the limit of poor modelocking (CW operation and CW with a sinusoidal ripple superimposed) through the overlapping pulse regime to the limit of well separated pulses, in which case the dnoidal function becomes a sequence of hyperbolic secants. It has been suggested that an approximate solution of the transient evolution problem in laser modelocking could be pieced together assuming that the pulse shape were approximately dnoidal at each point in time along the buildup (see Ref. 3).

Of central interest in the investigation of these simple models is the question of pulse stability. For example, one would expect that in a passively modelocked system that a perturbation introduced into the cavity which added in phase to the steady state pulse would be unstable for the following reason. Increased pulse amplitude causes additional absorber bleaching thereby decreasing the net loss seen by the pulse, causing it to grow. In fact, the system has a fundamental instability built into it through the nonlinear absorber saturation which can be stabilized only through an additional mechanism such as gain saturation.

In this communication we are concerned with the solution of the eigenvalue equation which plays a key role in the analysis of pulse stability in passively modelocked systems. The equation is a second order integro–differential equation which upon normalization, takes the following form,

$$su(x)=\left(\frac{d^2}{dx^2}-[(2-m)-6\mathrm{dn}^2(x,m)]\right)u(x)$$
$$-\epsilon\int_{-K(m)}^{K(m)}u(x')\,\mathrm{dn}(x',m)\,dx'\left(1+\nu^2\frac{d^2}{dx^2}\right)\mathrm{dn}(x,m), \tag{1.1}$$

where s is the eigenvalue and ϵ and ν are arbitrary parameters. $\mathrm{dn}(x,m)$ is the Jacobian elliptic dnoidal function of modulus m (see Fig. 1) and $K(m)$ is the complete elliptic integral of the first kind. The eigenfunctions $u(x)$ satisfy periodic boundary conditions

$$u(x+2K(m))=u(x) \tag{1.2}$$

which are the same boundary conditions satisfied by $\mathrm{dn}(x,m)$.

The eigenvalue s is proportional to the rate of growth of perturbations of the steady state as it passes back and forth within the laser cavity. The primary interest here is focussed on conditions on ϵ and ν under which there are no positive eigenvalues s. Physically, ϵ is a measure of the magnitude of the effect of gain saturation which is required for pulse stability as discussed earlier. The parameter ν is the ratio of the steady state pulse bandwidth and the gain linewidth, and is much less than unity wherever the model is valid. Our treatment of the equation, however, is valid for arbitrary ν. When ϵ is zero, the lowest eigenfunction of (1.1) has a positive eigenvalue s_0, and as the gain saturation parameter increases (for small ν), s_0 decreases monotonically until some critical value of ϵ there are no longer any positive eigenvalues. The result of the analysis of the following sections is the determination of the parameter ϵ as a function of the modulus m at which s_0 is zero.

We now review briefly some results which have been obtained on this equation which are relevant. When ϵ is zero, (1.1) reduces to a Lamé equation

$$s\Lambda(x)=\left(\frac{d^2}{dx^2}-[(2-m)-6\mathrm{dn}^2(x,m)]\right)\Lambda(x) \tag{1.3}$$

which has been investigated by Ince[4] and Erdelyi,[5]

[a)]Work performed in part under the auspices of the U. S. Department of Energy, Contract No. W-7405-Eng-48.
[b)]Work at MIT supported in part by the Fannie and John Hertz Foundation.

 0022-2488/78/1906-1386$1.00

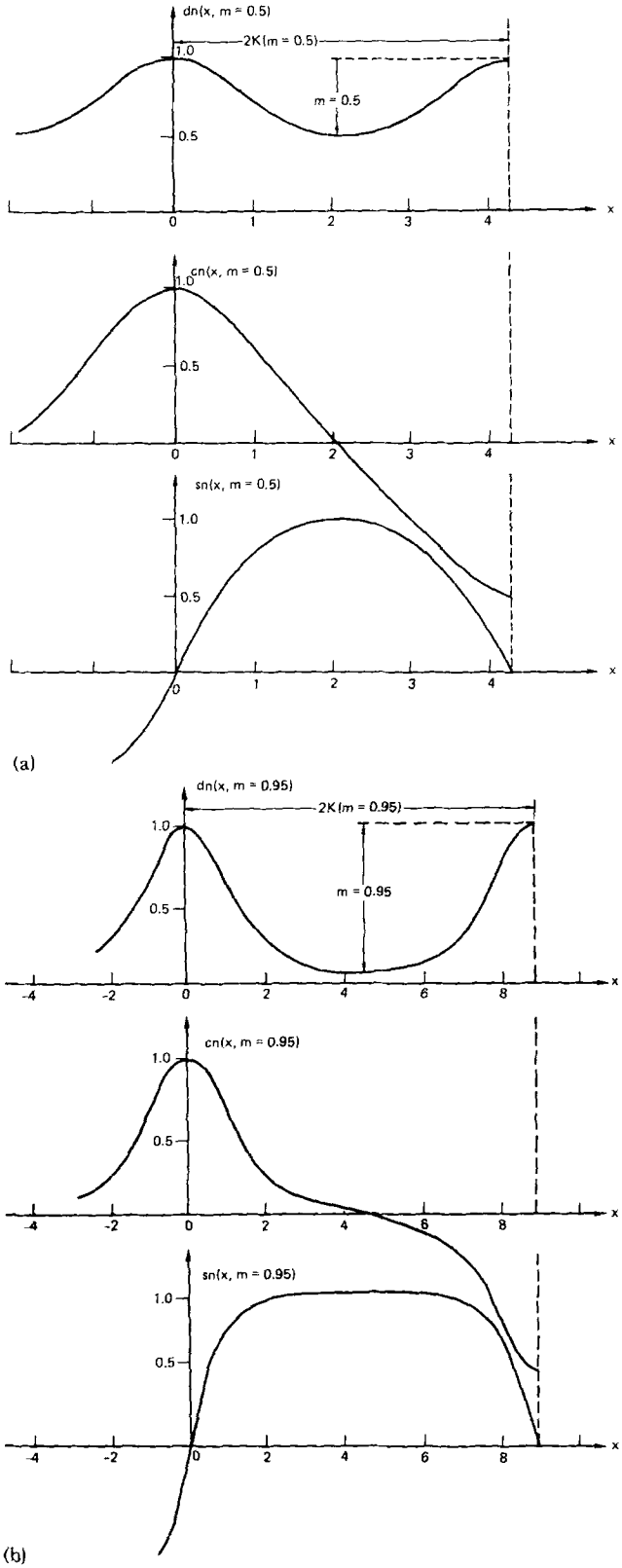


FIG. 1. Jacobian elliptic functions (a) $m=0.5$, (b) $m=0.95$.

The lowest two eigenfunctions are

$$\Lambda_0(x)=\mathrm{dn}^2(x,m)+\frac{(1-m+m^2)^{1/2}-(2-m)}{3} \tag{1.4}$$

$$s_0=(2-m)+2(1-m+m^2)^{1/2}$$

and

$$\Lambda_1(x)=\mathrm{sn}(x,m)\,\mathrm{cn}(x,m), \quad s_1=0, \tag{1.5}$$

where $\mathrm{cn}(x,m)$ and $\mathrm{sn}(x,m)$ are the Jacobian elliptic cnoidal and snoidal functions (see Fig. 1). As ϵ becomes finite and increases, we find that $\Lambda_1(x)$ is still a solution to (1.1) with eigenvalues $s_1=0$. Physically, this corresponds to a timing shift of the steady state pulse for which there is no stabilizing mechanism. It is the lowest even eigenfunction $u_0(x)$ which causes amplitude instabilities, and for this reason we shall focus on the construction of $u_0(x)$ in the following section.

When the pulses are well separated, that is when the modulus m approaches unity, (1.1) reduces to

$$su(x)=\left(\frac{d^2}{dx^2}-[1-6\ \mathrm{sech}^2x]\right)u(x)$$

$$-\epsilon\int_{-\infty}^{\infty}u(x')\,\mathrm{sech}(x')\,dx'\left(1+\nu^2\frac{d^2}{dx^2}\right)\mathrm{sech}(x). \tag{1.6}$$

This equation has been solved in a companion paper[6] with the result that

$$\epsilon=\frac{2}{1-\nu^2} \tag{1.7}$$

marks the stability boundary.

In the following sections we construct $u_0(x)$ from an expansion of the form

$$u_0(x)=\mathrm{dn}(x,m)\Big[\sum_{n\ \mathrm{even}}^{\infty}A_n\,\mathrm{cn}(nx,m)\Big]. \tag{1.8}$$

The recursion relations derived for the coefficients however are valid for all even eigenfunctions. This expansion was one form of expansion used by Ince in the solution of the Lamé equation[4] and is very convenient here because it leads to a simple three term recursion relation for the coefficients A_n, except in the case of A_0 and A_2 where one must also evaluate a rapidly convergent series. When s_0 is zero, one finds that

$$\epsilon=\frac{2}{\xi(m)-\nu^2[2E(m)-(2-m)\xi(m)]} \tag{1.9}$$

marks the stability boundary, where $E(m)$ is the complete elliptic integral of the second kind and where $\xi(m)$ is a function defined in Sec. IV.

II. THE RECURSION RELATION

In this section we derive the three term recursion relation for the coefficients A_n of the expansion of (1.8). We follow Ince[4] and transform (1.1) into a more convenient form using the substitutions

$$u(x)=\mathrm{dn}(x,m)\,f(x), \tag{2.1}$$

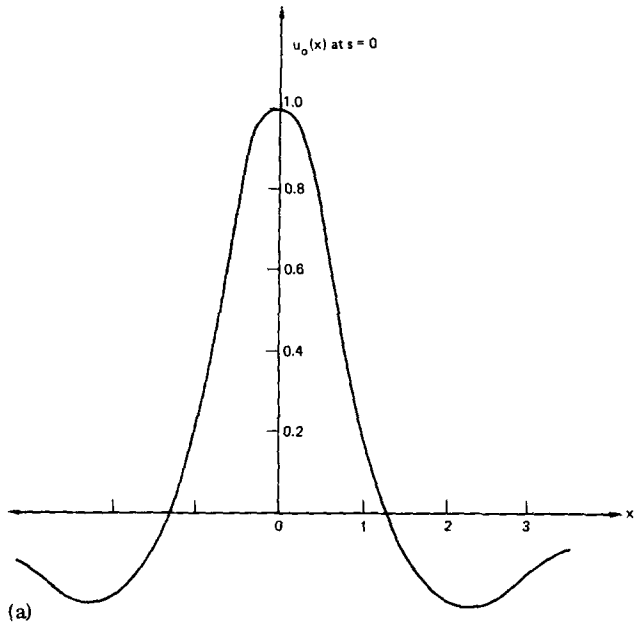
$$\cos(\theta)=\mathrm{cn}(x,m), \tag{2.2}$$

which yields

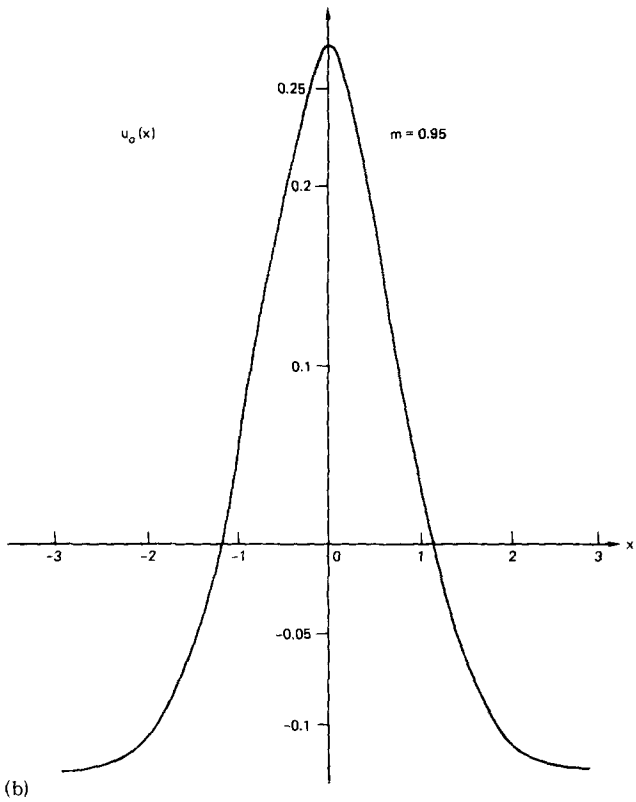
$$[(2-m)+m\cos 2\theta]\frac{d^2}{d\theta^2}f(\theta)-3m\sin\theta\frac{d}{d\theta}f(\theta)$$

$$+[(8-4m-2s)+4m\cos 2\theta]f(\theta)$$

$$-2\epsilon\alpha(1-m\nu^2\cos 2\theta)=0 \tag{2.3}$$



(a)



(b)

FIG. 2. (a) Lowest eigenfunction $u(x)$ at stability boundary ($s=0$) for $m=1$ and $\nu^2=0$. (b) Lowest even eigenfunction $u(x)$ at the stability boundary ($s=0$) for $m=0.95$.

where we have defined α to be the overlap integral

$$\alpha = \int_{-K(m)}^{K(m)} u(x)\,\mathrm{dn}(x,m)\,dx$$

$$= \int_{-\pi/2}^{\pi/2} f(\theta)\,(1 - m\sin^2\theta)^{1/2}\,d\theta. \tag{2.4}$$

Under the substitutions (2.1) and (2.2) the expansion of (1.8) becomes a Fourier series

$$f(\theta) = \sum_{n\ \text{even}} A_n \cos(n\theta) \tag{2.5}$$

which upon substitution into (2.3) leads to the following recursion relations:

$$\begin{aligned}
&\frac{3m}{2} A_2 + (4 - 2m - s)A_0 = \epsilon\alpha,\\
&-s A_2 + 2m A_0 = -\epsilon\alpha\, m\nu^2,\\
&\frac{m}{2}(2-n)(3+n)A_{n+2} + [(4-n^2)(2-m) - 2s]A_n\\
&+ \frac{m}{2}(2+n)(3-n)A_{n-2} = 0 \quad (n \geq 4).
\end{aligned} \tag{2.6}$$

Since we are primarily concerned with solutions of (1.1) when $s=0$, we consider solutions of (2.6) in this limit. We find that (2.6) becomes

$$\begin{aligned}
A_0 &= -\epsilon\alpha\,\frac{\nu^2}{2},\\
A_2 &= \frac{2\epsilon\alpha}{3m}[1 + \nu^2(2-m)]\\
&\quad \times \frac{m}{2}(2-n)(3+n)A_{n+2} + (4-n^2)(2-m)A_n\\
&\quad + \frac{m}{2}(2+n)(3-n)A_{n-2}\\
&= 0 \quad (n \geq 4)
\end{aligned} \tag{2.7}$$

when the modulus m is not zero. Asymptotically, A_n approaches zero in the following way,

$$A_n \underset{n\to\infty}{\longrightarrow} \left(\frac{-(2-m) + 2\sqrt{1-m}}{m}\right)^{2n} \tag{2.8}$$

for m less than unity. The convergence of the expansion is slow as m approaches unity and for m equal to unity (2.5) diverges at $\theta = 0$. This behavior is not unexpected since when at $m=1$, (2.5) becomes

$$u(x) = \mathrm{sech}\,x\left[\sum_{n\ \text{even}} A_n \,\mathrm{sech}(nx)\right] \tag{2.9}$$

which is an unfortunate choice of an expansion to fit a function like $u(x)$ as shown in Fig. 2(a). Terms of higher order in n become more and more spikelike around $x=0$ and provide poor higher order corrections to the series.

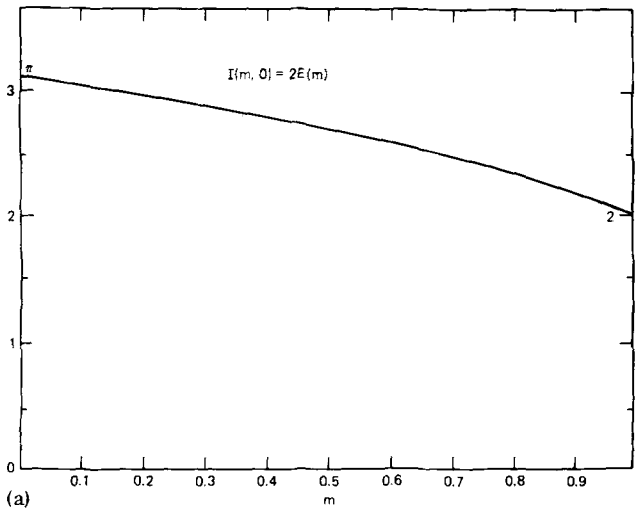
In Table I we give solutions to (2.7) for ν^2 equal to zero illustrating that only the first few terms of the expansion are significant for $m < 1$. In Fig. 2(b) we illustrate the lowest eigenfunction $u(x)$ at the stability boundary (at $s=0$) for $m=0.95$ in the case of $\nu^2=0$.

III. THE OVERLAP INTEGRAL

Ultimately we must solve for the overlap integral

TABLE I. Expansion coefficients at the stability boundary.

n	A_n $(m=0.5)$	A_n $(m=0.95)$
0	0	0
2	0.666	0.351
4	−0.0570	−0.105
6	0.00732	0.0483
8	−0.00105	−0.0250
10	0.000157	0.0137
12	−0.0000240	−0.00775



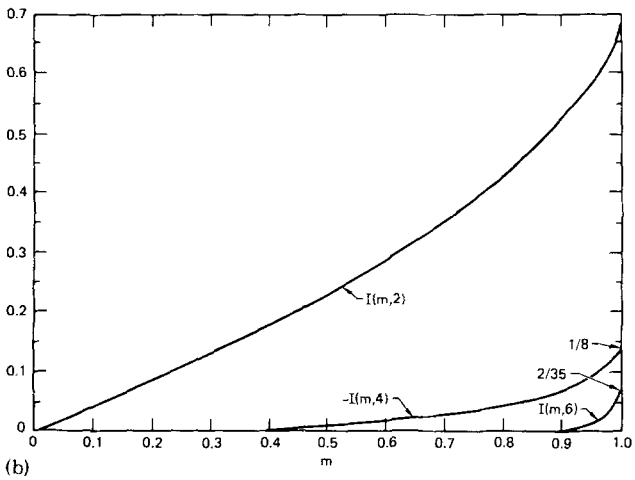


FIG. 3. $I(m,n)$ functions defined in (3.2). (a) $I(m,0)$; (b) $I(m,2)$, $-I(m,4)$, $I(m,6)$.

α self-consistently with the coefficients A_n. The integral of (2.4) can be written as

$$\alpha = \int_{-\pi/2}^{\pi/2} [{}_n\sum_{\text{even}} A_n \cos(n\theta)](1 - m\sin^2\theta)^{1/2}\, d\theta$$
$$= {}_n\sum_{\text{even}} A_n I(m,n), \tag{3.1}$$

where we define $I(m,n)$ to be

$$I(m,n) = \int_{-\pi/2}^{\pi/2} (1 - m\sin^2\theta)^{1/2} \cos(n\theta)\, d\theta. \tag{3.2}$$

Some of the lowest integrals $I(m,n)$ are shown in Fig. 3. One observes that (3.1) converges for all $0 \le m \le 1$ since in the worst case for A_n at $m = 1$,

$$I(1,n) = \frac{2(-1)^{n/2}}{1-n^2} \tag{3.3}$$

for even n.

We find that the series in (3.1) converges rapidly, especially for small values of the modulus m. The lowest order integral $I(m,0)$ can be written

$$I(m,0) = 2E(m), \tag{3.4}$$

where $E(m)$ is the complete elliptic integral of the second kind. For $n \ge 2$ we find

$$I(m,2) = m\frac{\pi}{2^3} + m^2\frac{5\pi}{2^7} + m^3\frac{2\cdot 7\pi}{2^{11}} + m^4\frac{5\cdot 7\cdot 9\pi}{2^{15}} + \cdots,$$
$$I(m,4) = -m^2\frac{\pi}{2^7} - m^3\frac{3\pi}{2^9} - m^4\frac{5\cdot 7\pi}{2^{13}} - \cdots, \tag{3.5}$$
$$I(m,6) = m^3\frac{\pi}{2^{10}} + m^4\frac{5\pi}{2^{12}} + \cdots.$$

IV. SOLUTION FOR THE PARAMETER ϵ AT THE STABILITY BOUNDARY

In this section we solve for the gain saturation parameter ϵ at the stability boundary, that is, when the eigenvalue s is zero. We begin by combining the recursion relations (2.7) with the expression for the overlap integral α (3.1) to give

$$\alpha = \frac{3mA_2}{2\epsilon}\,\frac{1}{[1+(2-m)\nu^2]} = {}_n\sum_{\text{even}} A_n I(m,n). \tag{4.1}$$

One can now solve this expression for the parameter ϵ to yield

$$\epsilon = \frac{3m}{2}\,\frac{1}{[1+(2-m)\nu^2]}\,\frac{1}{{}_n\sum_{\text{even}}^{\infty}(A_n/A_2)I(m,n)} \tag{4.2}$$

We note that

$$\frac{A_0}{A_2} = -\frac{3m\nu^2}{4[1+(2-m)\nu^2]} \tag{4.3}$$

which allows (4.2) to be written in the form

$$\epsilon = \frac{2}{\xi(m) - \nu^2[I(m,0) - (2-m)\xi(m)]}, \tag{4.4}$$

where we have defined the function $\xi(m)$ to be

$$\xi(m) = \frac{4}{3m}\sum_{\substack{n=2\\(n\text{ even})}}^{\infty}\frac{A_n}{A_2} I(m,n) \tag{4.5}$$

which is shown in Fig. 4. Noting that the function $I(0,m)$ can be written in terms of $E(m)$ through (3.4) yields

$$\epsilon = \frac{2}{\xi(m) - \nu^2[2E(m) - (2-m)\xi(m)]} \tag{4.6}$$

which is the required result.

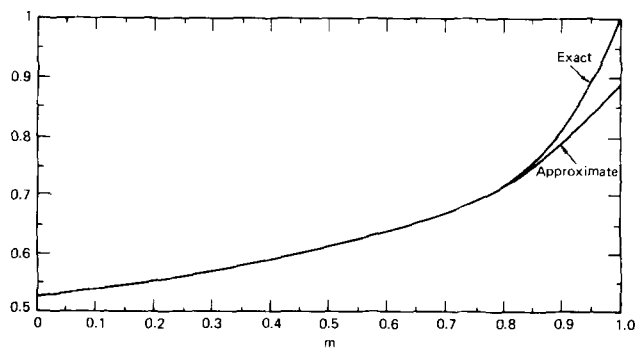


FIG. 4. The function $\xi(m)$ defined in (4.5) and the approximation (4.7).

It was found earlier that the coefficients A_n decrease rapidly with increasing order n, especially for small m, so that one might be tempted to approximate the series in (4.5) with the first few terms. For example, if one truncates the series after the first term, then one obtains

$$\xi(m) \approx \frac{4}{3m} I(m, 2) \tag{4.7}$$

which is exact for small m and turns out to be reasonable all the way to $m=1$ as is shown in Fig. 4. At the worse case of $m=1$, (4.7) underestimates the exact result by 11.1%.

V. SUMMARY

We have solved the eigenvalue equation

$$su(x) = \left(\frac{d^2}{dx^2} - [(2-m) - 6\mathrm{dn}^2(x, m)]\right)u(x) - \epsilon \int_{-K(m)}^{K(m)} u(x')\,\mathrm{dn}(x', m)\,dx' \left(1 + \nu^2 \frac{d^2}{dx^2}\right)\mathrm{dn}(x, m) \tag{5.1}$$

and have determined conditions on the parameter ϵ for which no eigenfunction $u(x)$ has a positive eigenvalue s for $u(x)$ which has a period $2K(m)$. The stability boundary where the lowest even eigenfunction has eigenvalue zero occurs when

$$\epsilon = \frac{2}{\xi(m) - \nu^2[2E(m) - (2-m)\xi(m)]} \tag{5.2}$$

where $\xi(m)$ is defined in (4.5) and is shown in Fig. 4 and where $E(m)$ is the elliptic integral of the second kind. We have found that the series expansion for $\xi(m)$ can be truncated after one term to yield the approximation

$$\xi(m) \approx \frac{4}{3m} I(m, 2) = \frac{4}{3m} \int_{-\pi/2}^{\pi/2} (1 - m\sin^2\theta)^{1/2} \cos(2\theta)\,d\theta, \tag{5.3}$$

which becomes exact for small m and underestimates the exact result at m equal to unity by 11.1%.

ACKNOWLEDGMENTS

The author takes pleasure in acknowledging stimulating discussions with Professor H. A. Haus at MIT. The work was supported at the Research Laboratory of Electronics by the Joint Services Electronics Program, Contract DAAB76-C-1400. P. L. Hagelstein was a Hertz Foundation Fellow. Use of MACSYMA Program (Mathlab Group, Project MAC, MIT) was greatly appreciated. The work of the Mathlab group is supported by ARPA Work Order 2095 under ONR Contract N00014-75-C-0661.

[1] H. A. Haus, "Theory of modelocking with a fast saturable absorber," in J. Appl. Phys. **46**, 3049 (1975).
[2] P. L. Hagelstein, "Laser modelocking with a saturable absorber," B. S. thesis MIT, Department of Electrical Engineering, March 1976, and "Stability analysis of a passively modelocked laser system," M. S. thesis, MIT, Department of Electrical Engineering, May 1976.
[3] C. P. Ausschnitt, "Transient evolution of passive modelocking," IEEE J. Quant. Elec. **QE-13**, 321 (1977), and "Transient evolution of passive modelocking—theory and experiment," Ph. D. Thesis, MIT, Department of Electrical Engineering, January 1976.
[4] E. Ince, "The periodic Lame functions," Proc. Roy. Soc. Edin. **60**, 47 (1940), and "Further investigations into the periodic Lame functions," Proc. Roy. Soc. Edin. **60**, 83 (1940).
[5] A. Erdelyi, *Higher Transcendental Functions, Vol. III* (McGraw-Hill, New York, 1955).
[6] P. L. Hagelstein, "Solution of a second-order integro—differential equation which occurs in laser modelocking," J. Math. Phys. (to be published).

The intelligent states. I. Group-theoretic study and the computation of matrix elements

M. A. Rashid[a)]

International Centre for Theoretical Physics, Trieste, Italy
(Received 4 August 1977)

In this first of a series of papers, a group-theoretic study is presented of the quasi-intelligent states which are a generalization of the intelligent states satisfying equality in the Heisenberg uncertainty relation $\Delta J_1^2 \Delta J_2^2 \geq (1/4)|\langle J_3\rangle|^2$. A method based on the knowledge of a certain generating function is given for the calculation of matrix elements of polynomials in the infinitesimal generators of the rotation group between quasi-intelligent states. Examples of such computations are also included to exhibit the improvement and efficiency of the present methods.

I. INTRODUCTION

Aragone *et al.*[1] have recently considered the states of a well-defined angular momentum which satisfy equality $\Delta J_1^2 \Delta J_2^2 = \frac{1}{4}|\langle J_3\rangle|^2$ in the Heisenberg uncertainty relation derived from the commutation relation $[J_1, J_2] = iJ_3$. Unlike the Glauber coherent states of a linear harmonic oscillator, these states are not generally the minimum uncertainty states, i.e., $\Delta J_1^2 \Delta J_2^2$ does not take a minimum value for them.

In the following series of papers, we shall attempt to present a somewhat different but more manifest method of handling these states which are called intelligent states in the literature. Our papers will clarify the algebraic structure of these states and emphasize the distinction between them and the usual Wigner states $|jm\rangle$.

In the present paper, we introduce the group-theoretic structure and present methods for the computation of elementary matrix elements of the generators between these states. In the second paper of the series we examine the problem of the computation of the Clebsch–Gordan coefficients for the intelligent states. In the third paper, we hope to present certain physical applications.

The present paper is organized as follows. In Sec. II, we repeat briefly, for completeness, the argument that the states which satisfy equality in the Heisenberg uncertainty relation are indeed eigenstates of a *non-Hermitian operator* J_3' with a known spectrum. In Sec. III, we give the operators which together with J_3' form the *same* algebra as that formed by the infinitesimal generators of the three-dimensional rotation group. In this section, we also present a compact representation of these states up to normalization in terms of the operation of the infinitesimal generators of the rotation group on the Wigner states. A simple expression for the normalization coefficients is also obtained in this section.

In Sec. IV, we arrive at a manifest connection between the intelligent states and the Wigner states. This connection also leads to another, somewhat more complicated, expression for the normalization coefficients which is shown to be equivalent to the simpler expression presented in Sec. III.

Section V is devoted to the computation of some elementary matrix elements by mentioning that a certain generating function is trivially calculable using our methods. With the use of this generating function, these elementary matrix elements can easily be calculated. This section ends with a few examples to illustrate the efficiency of our approach.

II. THE HEISENBERG UNCERTAINTY RELATION

Let us start with the commutator $[A, B] = iC$, where A and B are Hermitian (and hence C is also Hermitian). For any state ψ, defining

$$(\Delta A)^2 = \int \psi^* (A - \langle A\rangle)^2 \psi \, d\tau, \tag{1}$$

we obtain

$$(\Delta A)^2 = \int |(A - \langle A\rangle)|^2 \, d\tau \tag{2}$$

since A is Hermitian.

Now we use the Schwartz inequality

$$\int |f|^2 d\tau \int |g|^2 d\tau \geq \int |f^* g|^2 d\tau \geq \left| \int f^* g \, d\tau \right|, \tag{3}$$

with $f = (A - \langle A\rangle)\psi$ and $g = (B - \langle B\rangle)\psi$. This results in

$$(\Delta A)^2 (\Delta B)^2 \geq \left| \int \psi^* (A - \langle A\rangle)(B - \langle B\rangle) \psi \, d\tau \right|^2, \tag{4}$$

where the *equality* sign will hold if and only if

$$(B - \langle B\rangle)\psi = \lambda (A - \langle A\rangle)\psi, \tag{5}$$

where, so far, λ is any (possibly a complex) number.

Next we try to relate the right-hand side in the inequality given in Eq. (4) above to $\langle C\rangle$. We note that

$$\begin{aligned}(A - \langle A\rangle)(B - \langle B\rangle) &= \tfrac{1}{2}[(A - \langle A\rangle)(B - \langle B\rangle) + (B - \langle B\rangle)(A - \langle A\rangle)] \\ &\quad + \tfrac{1}{2}[(A - \langle A\rangle)(B - \langle B\rangle) - (B - \langle B\rangle)(A - \langle A\rangle)] \\ &= F + \tfrac{1}{2} iC, \end{aligned} \tag{6}$$

where

$$F = \tfrac{1}{2}[(A - \langle A\rangle)(B - \langle B\rangle) + (B - \langle B\rangle)(A - \langle A\rangle)] \tag{7a}$$

and

$$(A - \langle A\rangle)(B - \langle B\rangle) - (B - \langle B\rangle)(A - \langle A\rangle) = [A, B] = iC. \tag{7b}$$

[a)]On leave of absence from Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria.

 0022-2488/78/1906-1391$1.00

Since F and C are Hermitian operators, $\langle F\rangle$ and $\langle C\rangle$ are both real numbers and we find

$$|\langle F+\tfrac{1}{2}iC\rangle|^2=|\langle F\rangle|^2+\tfrac{1}{4}|\langle C\rangle|^2\geq\tfrac{1}{4}|\langle C\rangle|^2, \tag{8}$$

where again the equality will hold provided $\langle F\rangle=0$.

Combining Eqs. (4), (6), and (8), we arrive at

$$(\Delta A)^2(\Delta B)^2\geq\tfrac{1}{4}|\langle C\rangle|^2, \tag{9}$$

which is the well-known Heisenberg uncertainty relation.[2]

Our interest is basically in understanding when we shall have an equality in Eq. (9). From the argument presented above, it is now clear that the equality will hold for those states ψ for which

$$(B-\langle B\rangle)\psi=\lambda(A-\langle A\rangle)\psi \tag{10a}$$

and

$$\langle F\rangle=0. \tag{10b}$$

Equations (7) and (10) now lead to

$$\lambda(\Delta A)^2+\frac{1}{\lambda}(\Delta B)^2=0 \tag{11a}$$

and

$$\lambda(\Delta A)^2-\frac{1}{\lambda}(\Delta B)^2=i\langle C\rangle, \tag{11b}$$

which imply

$$\lambda=\tfrac{1}{2}i\,\frac{\langle C\rangle}{(\Delta A)^2}, \tag{12}$$

where since $\langle C\rangle$ and $(\Delta A)^2$ are both real, λ is indeed pure imaginary. This shows that the states ψ for which the Heisenberg uncertainty relation has an equality are those for which

$$(A-i\alpha B)\psi=(\langle A\rangle-i\alpha\langle B\rangle)\psi, \tag{13}$$

i.e., they are eigenstates of the operator $A-i\alpha B$ for *real* α. (Note that we have replaced the purely imaginary number λ^{-1} by $i\alpha$.)

Let us now apply the above result to the special case where A, B, C are J_1, J_2, J_3—the generators of the infinitesimal rotations in the three-dimensional space. Then we note that the states for which

$$(\Delta J_1)^2(\Delta J_2)^2=\tfrac{1}{4}|\langle J_3\rangle|^2$$

are eigenstates of the non-Hermitian operator $J_1-i\alpha J_2$ for some real α. In the following, we shall explicitly determine these states (called the intelligent states in the literature) for a given angular momentum j as a linear combination of the Wigner states $|jm\rangle$ and also study their properties.

III. THE OPERATORS $J_3'(\alpha)$ AND $J_\pm'(\alpha)$

Though intelligent states correspond to real α only, we shall consider the more general situation, where α is any complex number (the corresponding eigenstates of $J_1-i\alpha J_2$ may be called quasi-intelligent states).

We define[3]

$$J_3'(\alpha)=\frac{J_1-i\alpha J_2}{(1-\alpha^2)^{1/2}} \tag{14a}$$

and

$$J_\pm'(\alpha)=\mp\frac{\alpha}{(1-\alpha^2)^{1/2}}J_1\pm\frac{i}{(1-\alpha^2)^{1/2}}J_2-J_3 \tag{14b}$$

for any complex $\alpha\neq\pm1$. (This restriction will be clear soon.) The operators $J_3'(\alpha)$ and $J_\pm'(\alpha)$ satisfy the commutation relations

$$[J_3'(\alpha),J_\pm'(\alpha)]=\pm J_\pm'(\alpha) \tag{15a}$$

and

$$[J_+'(\alpha),J_-'(\alpha)]=2J_3'(\alpha), \tag{15b}$$

which are exactly the same as those satisfied by $J_3, J_\pm=J_1\pm iJ_2$. Also

$$\begin{aligned}\mathbf{J}^2&=J_1^2+J_2^2+J_3^2=\tfrac{1}{2}(J_+J_-+J_-J_+)+J_3^2\\&=\tfrac{1}{2}[J_+'(\alpha)J_-'(\alpha)+J_-'(\alpha)J_+'(\alpha)]+J_3'^2(\alpha).\end{aligned} \tag{16}$$

We try to construct the eigenstates of the operator $J_3'(\alpha)$ as a linear combination of the states $|jm\rangle$ with a given j and $-j\leq m\leq j$, where these $(2j+1)$ Wigner states are the eigenstates of the *Hermitian operator* J_3. Indeed

$$J_3|jm\rangle=m|jm\rangle \tag{17a}$$

and

$$J_\pm|jm\rangle=\sqrt{(j-m)(j+m+1)}\,|jm\rangle. \tag{17b}$$

Since J_3 and $\mathbf{J}^2$ are *Hermitian operators*, the states $|jm\rangle$ can be orthonormalized in the form

$$\langle j'm'|jm\rangle=\delta_{jj'}\delta_{mm'}, \tag{18}$$

which is what one conventionally does.

Noting that

$$\exp(J_3\theta)J_1\exp(-J_3\theta)=J_1\cosh\theta+iJ_2\sinh\theta, \tag{19}$$

we realize that the right-hand side will be proportional to $J_1-i\alpha J_2$ provided one chooses θ such that

$$\cosh\theta=\frac{1}{(1-\alpha^2)^{1/2}}\quad\text{and}\quad\sinh\theta=-\frac{\alpha}{(1-\alpha^2)^{1/2}} \tag{20}$$

or

$$e^\theta=\left(\frac{1-\alpha}{1+\alpha}\right)^{1/2}=\tau \tag{21}$$

and then the right-hand side of Eq. (19) is just $J_3'(\alpha)$. With the above choice of θ, we find

$$\exp(\pm J_3\theta)J_1\exp(\mp J_3\theta)=\frac{J_1\mp i\alpha J_2}{(1-\alpha^2)^{1/2}} \tag{22a}$$

and

$$\exp(\pm J_3\theta)J_2\exp(\mp J_3\theta)=\frac{\pm i\alpha J_1+J_2}{(1-\alpha^2)^{1/2}}. \tag{22b}$$

In particular,

$$\exp(-J_3\theta)J_3'(\alpha)\exp(J_3\theta)=J_1. \tag{23}$$

Noting also that

$$\exp(\pm i\tfrac{1}{2}\pi J_2)J_1\exp(\mp i\tfrac{1}{2}\pi J_2)=\pm J_3, \tag{24}$$

we immediately see that the state

$$|jm\alpha\rangle'=\exp(\theta J_3)\exp(-i\tfrac{1}{2}\pi J_2)|jm\rangle \tag{25}$$

is indeed an eigenstate of the operator $J_3'(\alpha)$ with the eigenvalue m. The prime on $|jm\alpha\rangle$ is indicative of the fact that the state as defined may not be normalized.

Since θ is not necessarily pure imaginary (this requires $|\tau|=1$ or $\alpha=$ pure imaginary—for intelligent states θ is definitely not pure imaginary[4]) the above state does not, in general, correspond to a rotation of the Wigner state $|jm\rangle$. In fact, as we shall see, the above state is not normalized and the $(2j+1)$ states $|jm\alpha\rangle'$ for $-j\leq m\leq j$ are not orthogonal unless θ is purely imaginary. *The basic reason for this is the non-Hermiticity of the operator $J_3'(\alpha)$ of which these are eigenstates.*

Next we attempt to compute the overlap

$$
\begin{aligned}
&'\langle jm'\alpha'|jm\alpha\rangle' \\
&\quad=\langle jm'|\exp(i\tfrac{1}{2}\pi J_2)\exp(\theta'^{*}J_3)\exp(\theta J_3)\exp(-i\tfrac{1}{2}\pi J_2)|jm\rangle \\
&\quad=\langle jm'|\exp[-(\theta+\theta'^{*})J_1]|jm\rangle .
\end{aligned}
\tag{26}
$$

The above matrix element can immediately be computed using the 2×2 representation

$$J_1\cong\tfrac{1}{2}\sigma_1=\tfrac{1}{2}\begin{bmatrix}0&1\\1&0\end{bmatrix}$$

of J_1, in terms of which

$$
\exp[-(\theta+\theta'^{*})J_1]\cong\begin{bmatrix}\cosh\left(\frac{\theta+\theta'^{*}}{2}\right)&-\sinh\left(\frac{\theta+\theta'^{*}}{2}\right)\\-\sinh\left(\frac{\theta+\theta'^{*}}{2}\right)&\cosh\left(\frac{\theta+\theta'^{*}}{2}\right)\end{bmatrix}
\tag{27}
$$

Thus we find

$$
\begin{aligned}
&'\langle jm'\alpha'|jm\alpha\rangle' \\
&=\langle jm'|\exp[-(\theta+\theta'^{*})J_1]|jm\rangle \\
&=[(j+m)!(j-m)!(j+m')!(j-m')!]^{1/2}(-1)^{2j+m+m'} \\
&\quad\times\sum_r\frac{[\cosh((\theta+\theta'^{*})/2)]^{-m-m'+2r}[\sinh((\theta+\theta'^{*})/2)]^{2j+m+m'-2r}}{r!(j+m-r)!(j+m'-r)!(-m-m'+r)!}
\end{aligned}
\tag{28a}
$$

$$
\begin{aligned}
&=\left(\frac{(j+m)!(j-m')!}{(j-m)!(j+m')!}\right)^{1/2}\left[\sinh\left(\frac{\theta+\theta'^{*}}{2}\right)\right]^{m'-m}(-1)^{2j+m+m'} \\
&\quad\times\sum_r(-1)^r\frac{(2j-r)!\,[\cosh((\theta+\theta'^{*})/2)]^{2j+m-m'-2r}}{r!(j+m-r)!(j-m'-r)!}\,.
\end{aligned}
\tag{28b}
$$

In particular, if we define the normalized state by

$$
\begin{aligned}
|jm\alpha\rangle&=(a^j{}_m(\alpha))^{-1}|jm\alpha\rangle' \\
&=(a^j{}_m(\alpha))^{-1}\exp(\theta J_3)\exp(-i\tfrac{1}{2}\pi J_2)|jm\rangle,
\end{aligned}
\tag{29}
$$

then the normalization coefficient $a^j{}_m(\alpha)$ is given by

$$
\begin{aligned}
a^j{}_m(\alpha)&=\Bigg[(j+m)!(j-m)!\sum_r\frac{[\cosh((\theta+\theta^{*})/2)]^{-2m+2r}}{r!(j+m-r)!} \\
&\quad\times\frac{[\sinh((\theta+\theta^{*})/2)]^{2j+2m-2r}}{(j+m-r)!(-2m+r)!}\Bigg]^{1/2}.
\end{aligned}
\tag{30a}
$$

$$
=\left[\sum_r(-1)^r\frac{(2j-r)!\,[\cosh((\theta+\theta^{*})/2)]^{2j-2r}}{r!(j+m-r)!(j-m-r)!}\right]^{1/2}.
\tag{30b}
$$

In Eq. (30a), since $\sinh((\theta+\theta^{*})/2)$ is raised to an *even* power, every term in the sum is nonnegative ($\theta+\theta^{*}$ is a real quantity) and hence the square root is well defined.

Since the angle θ in everything we consider appears as a function of $\exp(\theta/2)=\tau^{1/2}$, we shall choose the τ complex plane to be cut from $-\infty$ to 0. The relationship between α and τ now is

$$\tau=\left(\frac{1-\alpha}{1+\alpha}\right)^{1/2}\quad\text{or}\quad\alpha=\frac{1-\tau^2}{1+\tau^2}.\tag{31a}$$

In particular

$$1-\alpha^2=\frac{4\tau^2}{(1+\tau^2)^2}\tag{31b}$$

and we choose

$$(1-\alpha^2)^{1/2}=\frac{2\tau}{1+\tau^2}.\tag{31c}$$

Thus any expression in terms of θ and α can immediately be expressed in terms of the variable τ. For real α (i.e., when we are considering intelligent states only), τ will be taken as positive real or positive imaginary. Indeed, all the α plane is obtained from the upper half-plane of τ.

We end this section with a few remarks.

(1) The matrix elements $\langle jm'|\exp(-\theta J_1)|jm\rangle$ are evidently [see Eq. (28a) with θ in place of $\theta+\theta'^{*}$] symmetrical in the interchange $m\rightleftharpoons m'$. For real θ, these are also real. The expression in Eq. (28b) is not manifestly symmetrical under this interchange, but it is presented on account of its simplicity and usefulness.

(2) The matrix elements $\langle jm'|\exp(-\theta J_1)|jm\rangle$ *for real* θ cannot be zero except when $\theta=0$ since every term on the right-hand side in Eq. (28a) is positive for $\theta<0$ and is positive (negative) for $\theta>0$ whenever $2j+m+m'$ is even (odd). This shows that the states $|jm\alpha\rangle$, $-j\leq m\leq j$, for given j and α are not orthogonal unless real $\theta=0$. The quasi-intelligent states for a given j and α are thus necessarily nonorthogonal unless $|\tau|=1$ or α pure imaginary.

(3) The normalization factors $a^j{}_m(\alpha)$ for a given j and α are only 1 when real $\theta=\frac{1}{2}(\theta+\theta^{*})=0$. Indeed whenever real $\theta=0$, $[a^j{}_m(\alpha)]^2=1$ as is obvious from Eq. (30a). For real $\theta\neq0$, we can differentiate the expression for $[a^j{}_m(\alpha)]^2$ obtained from Eq. (30a) and note that the derivative has the sign of $\tanh((\theta+\theta^{*})/2)$. Thus the norms $a^j{}_m(\alpha)\gtrless1$ for $\theta+\theta^{*}\gtrless0$.

Aragone *et al.*[5] could not find the properties mentioned in remarks (2) and (3) above, since they lacked simple analytic expressions for the matrix elements of the form $\langle jm'|\exp(-\theta J_1)|jm\rangle$.

(4) In the next section, we shall rewrite Eq. (29) expressing the states $|jm\alpha\rangle$ for any given m as a linear combination of the $(2j+1)$ Wigner states $|jm'\rangle$, $-j\leq m'\leq j$. This process can also be inverted, i.e., we can express any Wigner state $|jm\rangle$ in terms of the $(2j+1)$ quasi-intelligent states $|jm'\alpha\rangle$, $-j\leq m'\leq j$ for a given $\alpha\neq\pm1$. (This inversion will be presented in the second paper.) In this sense, therefore, the $(2j+1)$ quasi-intelligent states $|jm'\alpha\rangle$, $-j\leq m'\leq j$ for a given $\alpha\neq\pm1$ are complete. For $\alpha=+1(-1)$, $J_3'(\alpha)=J_1-iJ_2$

$(J_1 + iJ_2)$ is the usual lowering (raising) operator for the Wigner states. The only eigenstate it has is the Wigner state $|j, -j\rangle$ ($|j, j\rangle$) with the eigenvalue 0. Thus for $\alpha = \pm 1$, the analysis presented above completely breaks down. The question of completeness for $|jm\alpha\rangle$, $\alpha = \pm 1$ is just not there. This is the reason why in the beginning of the present section we restricted ourselves to $\alpha \neq \pm 1$.

IV. RELATIONSHIP BETWEEN THE QUASI-INTELLIGENT STATES AND THE WIGNER STATES

We have seen that the normalized quasi-intelligent states for given j, α are given by

$$|jm\alpha\rangle = (a^j{}_m(\alpha))^{-1} \exp(\theta J_3) \exp(-i\tfrac{1}{2}\pi J_2)\,|jm\rangle. \tag{29'}$$

Again we can use the 2×2 representation $J_i \cong \frac{1}{2}\sigma_i$ for the operators J_i [6] and employ the standard techniques to arrive at

$$|jm\alpha\rangle = (a^j{}_m(\alpha))^{-1}\, 2^{-j} \sum_{m'r} |jm'\rangle \exp(m'\theta)(-1)^{j+m'-\gamma} \times \frac{[(j+m)!\,(j-m)!\,(j+m')!\,(j-m')!\,]^{1/2}}{r!(j+m-r)!\,(j+m'-r)!\,(-m-m'+r)!} \tag{32a}$$

$$= (a^j{}_m(\alpha))^{-1}\, 2^{-j} \sum_{m'r} |jm'\rangle \exp(m'\theta)\, 2^r (-1)^r \times \left[\frac{(j-m)!\,(j+m')!}{(j+m)!\,(j-m')!}\right]^{1/2} \frac{(2j-r)!}{r!\,(j-m-r)!\,(j+m'-r)!} \tag{32b}$$

$$= (a^j{}_m(\alpha))^{-1}\, 2^{-j} \sum_{m'r} |jm'\rangle \exp(m'\theta)\, 2^r (-1)^{m'-m+r} \times \left[\frac{(j+m)!\,(j-m')!}{(j-m)!\,(j+m')!}\right]^{1/2} \frac{(2j-r)!}{r!\,(j+m-r)!\,(j-m'-r)!} \tag{32c}$$

Using the above results in various combinations, we obtain several equivalent expressions for the inner product $\langle jm'\alpha'|jm\alpha\rangle$. Thus, for example, from Eqs. (32b) and (32c), we find

$$\langle jm'\alpha'|jm\alpha\rangle = (a^j{}_m(\alpha)\, a^j{}_{m'}(\alpha'))^{-1} \left[\frac{(j+m)!\,(j-m')!}{(j-m)!\,(j+m')!}\right]^{1/2} \times \sum_{nrs} (-1)^{n-m+r+s}\, 2^{-2j+r+s} \exp[n(\theta+\theta'^*)] \times \frac{(2j-r)!\,(2j-s)!}{r!\,(j+m-r)!\,(j-n-r)!\,s!\,(j-m'-s)!\,(j+n-s)!}. \tag{33}$$

Though the above expression is not *manifestly* symmetrical under the interchange $m \rightleftharpoons m'$ (a symmetrical expression is obtained using the same representation for both $|jm\alpha\rangle$ and $|jm'\alpha'\rangle$), yet it is useful in establishing its relationship with the expression given earlier in Eq. (28b) when multiplied with $[a^j{}_m(\alpha)\, a^j{}_{m'}(\alpha')]^{-1}$. Indeed, we can perform the n-summation immediately. [Note that n in Eq. (33) need not be an integer though $j \pm n$, $m \pm n$ are so.] This allows us to rewrite the above results as

$$\langle jm'\alpha'|jm\alpha\rangle = (a^j{}_m(\alpha)\, a^j{}_{m'}(\alpha'))^{-1} \left[\frac{(j+m)!\,(j-m')!}{(j-m)!\,(j+m')!}\right]^{1/2} \times \sum_{rs} (-1)^{j+m-r} \exp[-(j-s)(\theta+\theta'^*)] \times [\tfrac{1}{2}(1-\exp(\theta+\theta'^*))]^{2j-r-s} \times \frac{(2j-r)!\,(2j-s)!}{r!\,(j+m-r)!\,s!\,(j-m'-s)!\,(2j-r-s)!}. \tag{34}$$

Using the transformation

$$F_2(\alpha, \beta, \beta', \alpha, \alpha; x, y) = (1-x)^{-\beta}(1-y)^{-\beta'}\, {}_2F_1\left(\beta, \beta; \alpha; \frac{xy}{(1-x)(1-y)}\right),$$

where

$$F_2(\alpha, \beta, \beta', \gamma, \gamma'; x, y) = \sum_{m,n} \frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!\,n!\,(\gamma)_m(\gamma')_n}\, x^m y^n$$

and ${}_2F_1$ is the usual hypergeometrical function,[7] we reproduce the expression for $\langle jm'\alpha'|jm\alpha\rangle$ mentioned above.

V. COMPUTATION OF THE MATRIX ELEMENTS BETWEEN QUASI-INTELLIGENT STATES

In this section, we present a method, essentially based on the knowledge of a generating function, for computing matrix elements of polynomial functions of the infinitesimal generators of the rotation group between quasi-intelligent states. We first define a generating function

$$G(j, m_1, m_2; \alpha_1, \alpha_2; a, b, c) = \langle jm_2\alpha_2|\, \exp\{c[J_3'(\alpha_2)]^\dagger\} \exp(bJ_3) \exp[aJ_3'(\alpha_1)]\,|jm_1\alpha_1\rangle. \tag{35}$$

Using

$$J_3'(\alpha)\,|jm\alpha\rangle = m\,|jm\alpha\rangle,$$

and its adjoint[8]

$$\langle jm\alpha|\,[J_3'(\alpha)]^\dagger = \langle jm\alpha|\, m,$$

we can rewrite the above generating function as

$$G(j, m_1, m_2; \alpha_1, \alpha_2; a, b, c) = \exp(am_1 + cm_2)\langle jm_2\alpha_2|\, \exp(bJ_3)\,|jm_1\alpha_1\rangle. \tag{36}$$

Now we use the method presented in Sec. III to arrive at [see Eq. (28)]

$$G(j, m_1, m_2; \alpha_1, \alpha_2; a, b, c) = \frac{\exp(am_1 + cm_2)}{a^j{}_{m_1}(\alpha_1)\, a^j{}_{m_2}(\alpha_2)} [(j+m_1)!\,(j-m_1)!\,(j+m_2)!\,(j-m_2)!]^{1/2} \times (-1)^{2j+m_1+m_2} \sum_r [\cosh\tfrac{1}{2}(\theta_1 + \theta_2^* + b)]^{-m_1-m_2+2r} \times [\sinh\tfrac{1}{2}(\theta_1 + \theta_2^* + b)]^{2j+m_1+m_2-2r}/r!\,(j+m_1-r)! \times (j+m_2-r)!\,(-m_1-m_2+r)! \tag{37a}$$

$$= \frac{\exp(am_1 + cm_2)}{a^j{}_{m_1}(\alpha_1)\, a^j{}_{m_2}(\alpha_2)} \left[\frac{(j+m_1)!\,(j-m_2)!}{(j-m_1)!\,(j+m_2)!}\right]^{1/2} \times [\sinh\tfrac{1}{2}(\theta_1 + \theta_2^* + b)]^{m_2-m_1} (-1)^{2j+m_1+m_2} \times \sum_r (-1)^r \frac{(2j-r)!\,[\cosh\tfrac{1}{2}(\theta_1 + \theta_2^* + b)]^{2j+m_1-m_2-2r}}{r!\,(j+m_1-r)!\,(j-m_2-r)!}. \tag{37b}$$

This generating function immediately gives the matrix element

$$\langle jm\alpha_2|[(J_3^\dagger(\alpha_2))^\dagger]^{n_2}[J_3]^n[J_3^\dagger(\alpha_1)]^{n_1}|jm_1\alpha_1\rangle$$
$$=\frac{m_1^{n_1}m_2^{n_2}}{a^j{}_{m_1}(\alpha_1)\,a^j{}_{m_2}(\alpha_2)}[(j+m_1)!\,(j-m_1)!\,(j+m_2)!\,(j-m_2)!]^{1/2}$$
$$\times(-1)^{2j+m_1+m_2}\frac{\partial^n}{\partial\theta_1^n}\Big[\sum_r[\cosh\tfrac{1}{2}(\theta_1+\theta_2^*)]^{-m_1-m_2+2r}$$
$$\times[\sinh\tfrac{1}{2}(\theta_1+\theta_2^*)]^{2j+m_1+m_2-2r}/r!\,(j+m_1-r)!$$
$$\times(j+m_2-r)!\,(-m_1-m_2+r)!\Big] \tag{38}$$

In particular,

$$\langle j_2m_2\alpha_2|J_3|j_1m_1\alpha_1\rangle$$
$$=\frac{[(j+m_1)!\,(j-m_1)!\,(j+m_2)!\,(j-m_2)!]^{1/2}}{a^j{}_{m_1}(\alpha_1)\,a^j{}_{m_2}(\alpha_2)}(-1)^{2j+m_1+m_2}$$
$$\times\sum_r\{[\cosh\tfrac{1}{2}(\theta_1+\theta_2^*)]^{-m_1-m_2+2r}$$
$$\times[\sinh\tfrac{1}{2}(\theta_1+\theta_2^*)]^{2j+m_1+m_2-2r}\}$$
$$\times[r!\,(j+m_1-r)!\,(j+m_2-r)!\,(-m_1-m_2+r)!\,]^{-1}$$
$$\times\tfrac{1}{2}[(-m_1-m_2+2r)\tanh\tfrac{1}{2}(\theta_1+\theta_2^*)$$
$$+(2j+m_1+m_2-2r)\coth\tfrac{1}{2}(\theta_1+\theta_2^*)]. \tag{39}$$

The diagonal matrix elements of J_3 are much more interesting. We find

$$\langle jm\alpha|J_3|jm\alpha\rangle$$
$$=-j\cos\delta\Big[1+\frac{1}{j}$$
$$\times\frac{\sum_r\sin^2\delta(\cos\delta)^{2r-2}/r!\,(r-1)!\,(j+m-r)!\,(j-m-r)!}{\sum_r(\cos\delta)^{2r}/(r!)^2(j+m-r)!\,(j-m-r)!}\Big], \tag{40}$$

where we have used the notation

$$\tan\tfrac{1}{2}\delta=|\tau|=|e^\theta| \tag{41a}$$

in terms of which

$$\tanh(\tfrac{1}{2}(\theta+\theta^*))=-\cos\delta, \tag{41b}$$
$$\cosh^2(\tfrac{1}{2}(\theta+\theta^*))=(\sin\delta)^{-2}. \tag{41c}$$

Equation (40) above is a generalization of the special results given in Eq. (40) of Ref. 1 with δ in our result corresponding to the θ in this reference. One has only to go through the two calculations to appreciate the simplicity and clarity of our methods. Also we have manifested the part which goes to zero when $\delta\to 0$, i.e., when $\alpha\to 1$ when m can only take the value $-j$ and the state $|jm\alpha\rangle$ can be none but the Wigner state $|j(-j)\rangle$. Incidentally this also shows that the results in Ref. 1 are wrong by a factor of 2.

Next we calculate the matrix elements of J_1 and J_2 between the quasi-intelligent states.

From

$$\langle jm_2\alpha_2|J_1-i\alpha_1J_2|jm_1\alpha_1\rangle=m_1(1-\alpha_1{}^2)^{1/2}=\frac{2m_1\tau_1}{1+\tau_1^2} \tag{42a}$$

and

$$\langle jm_2\alpha_2|J_1+i\alpha_2^*J_2|jm_1\alpha_1\rangle=m_2(1-\alpha_2{}^2)^{1/2*}=\frac{2m_2\tau_2^*}{1+\tau_2^{2*}}, \tag{42b}$$

where

$$\alpha_1=\frac{1-\tau_1^2}{1+\tau_1^2}\quad\text{and}\quad\alpha_2=\frac{1-\tau_2^2}{1+\tau_2^2},$$

we conclude

$$\langle jm_2\alpha_2|J_1|jm_1\alpha_1\rangle$$
$$=\frac{m_1\alpha_2^*(1-\alpha_1{}^2)^{1/2}+m_2\alpha_1(1-\alpha_2{}^2)^{1/2*}}{\alpha_1+\alpha_2^*} \tag{43a}$$
$$=\frac{m_1\tau_1(1-\tau_2^{2*})+m_2\tau_2^*(1-\tau_1^2)}{1-\tau_1^2\tau_2^{2*}} \tag{43b}$$

and

$$\langle jm_2\alpha_2|J_2|jm_1\alpha_1\rangle$$
$$=i\frac{m_1(1-\alpha_1{}^2)^{1/2}-m_2(1-\alpha_2{}^2)^{1/2*}}{\alpha_1+\alpha_2^*} \tag{44a}$$
$$=i\frac{m_1\tau_1(1+\tau_2^{2*})-m_2\tau_2^*(1+\tau_1^2)}{1-\tau_1^2\tau_2^{2*}}. \tag{44b}$$

In particular

$$\langle jm\alpha_2|J_1|jm\alpha_1\rangle=\frac{m(\tau_1+\tau_2^*)}{1+\tau_1\tau_2^*} \tag{45a}$$

and

$$\langle jm\alpha_2|J_2|jm\alpha_1\rangle=i\frac{m(\tau_1-\tau_2^*)}{1+\tau_1\tau_2^*}. \tag{45b}$$

If we also wish to consider the special case where $\alpha_1=\alpha_2$, we shall have

$$\langle jm\alpha|J_1|jm\alpha\rangle=2m\frac{\mathrm{Re}\tau}{1+|\tau|^2} \tag{46a}$$

and

$$\langle jm\alpha|J_2|jm\alpha\rangle=-2m\frac{\mathrm{Im}\tau}{1+|\tau|^2}. \tag{46b}$$

In Eq. (18) of Ref. 1 we note the factors $\mathrm{Re}\tau$ and $\mathrm{Im}\tau$. The remaining part of the above very simple answer is hidden in uncomputed derivatives of certain generating functions. With very little work, we have been able to compute even the nondiagonal matrix elements. We remark that since J_1 and J_2 are Hermitian, our matrix elements in Eq. (44) must be invariant under the combined operation of $m_1\alpha_1\rightarrow m_2\alpha_2$ and complex conjugation and indeed they are evidently so. The same property demands that the matrix elements in Eq. (46) be real and these are as expected.

As another example of our method, we now obtain the matrix elements of J_1^2, $J_1J_2+J_2J_1$, and J_2^2. We shall be able to express these matrix elements in terms of those of J_3 which have been given earlier in Eq. (39).

Starting from

$$\langle jm_2\alpha_2|(J_1-i\alpha_1J_2)^2|jm_1\alpha_1\rangle=m_1^2(1-\alpha_1^2), \tag{47a}$$
$$\langle jm_2\alpha_2|(J_1+i\alpha_2{}^*J_2)^2|jm_1\alpha_1\rangle=m_2^2(1-\alpha_2{}^2)^*, \tag{47b}$$

and

$$\langle jm_2\alpha_2|(J_1+i\alpha_2{}^*J_2)(J_1-i\alpha_1J_2)|jm_1\alpha_1\rangle$$
$$=m_1m_2(1-\alpha_1{}^2)^{1/2}(1-\alpha_2{}^2)^{1/2*} \tag{47c}$$

we arrive at

$$\langle jm_2\alpha_2|J_1^2|jm_1\alpha_1\rangle = \left[\frac{m_1(1-\alpha_1{}^2)^{1/2}\alpha_2{}^* + m_2\alpha_1(1-\alpha_2{}^2)^{1/2*}}{\alpha_1+\alpha_2{}^*}\right]^2 - \frac{\alpha_1\alpha_2{}^*}{\alpha_1+\alpha_2{}^*}\langle jm_2\alpha_2|J_3|jm_1\alpha_1\rangle, \tag{48a}$$

$$\langle jm_2\alpha_2|\tfrac{1}{2}(J_1J_2+J_2J_1)|jm_1\alpha_1\rangle = i[(m_1(1-\alpha_1{}^2)^{1/2} - m_2(1-\alpha_2{}^2)^{1/2*})(m_1\alpha_2{}^*(1-\alpha_1{}^2)^{1/2} + m_2\alpha_1(1-\alpha_2{}^2)^{1/2*})]/\alpha_1+\alpha_2{}^* - \frac{i}{2}\frac{\alpha_1-\alpha_2{}^*}{\alpha_1+\alpha_2{}^*}\langle jm_2\alpha_2|J_3|jm_1\alpha_1\rangle \tag{48b}$$

and

$$\langle jm_2\alpha_2|J_2{}^2|jm_1\alpha_1\rangle = -\left[\frac{m_1(1-\alpha_1{}^2)^{1/2} - m_2(1-\alpha_2{}^2)^{1/2*}}{\alpha_1+\alpha_2{}^*}\right]^2 - \frac{1}{\alpha_1+\alpha_2{}^*}\langle jm_2\alpha_2|J_3|jm_1\alpha_1\rangle. \tag{48c}$$

Note that the parts on the right-hand side of Eqs. (48) above, which are independent of J_3, could have been obtained from the matrix elements of J_1 and J_2 given earlier in Eqs. (43) and (44), as should be the case, since these are the values of the corresponding matrix elements *provided* J_1 and J_2 commuted. Also defining

$$\langle jm_2\alpha_2|\Delta J_i^2|jm_1\alpha_1\rangle = \langle jm_2\alpha_2|J_i^2|jm_1\alpha_1\rangle - (\langle jm_2\alpha_2|J_i|jm_1\alpha_1\rangle)^2 \tag{49}$$

for $i=1,2$, we obtain

$$\langle jm_2\alpha_2|\Delta J_1^2|jm_1\alpha_1\rangle = -\frac{\alpha_1\alpha_2{}^*}{\alpha_1+\alpha_2{}^*}\langle jm_2\alpha_2|J_3|jm_1\alpha_1\rangle \tag{50a}$$

and

$$\langle jm_2\alpha_2|\Delta J_2^2|jm_1\alpha_1\rangle = -\frac{1}{\alpha_1+\alpha_2{}^*}\langle jm_2\alpha_2|J_3|jm_1\alpha_1\rangle. \tag{50b}$$

In the above results, the m-dependence of the matrix elements on the left is *entirely* given in terms of the m-dependence of the matrix elements of J_3.

From Eqs. (50), we conclude

$$\langle jm_2\alpha_2)|\Delta J_1^2|jm_1\alpha_1\rangle\langle jm_2\alpha_2|\Delta J_2^2|jm_1\alpha_1\rangle = \frac{\alpha_1\alpha_2{}^*}{(\alpha_1+\alpha_2{}^*)^2}(\langle jm_2\alpha_2|J_3|jm_1\alpha_1\rangle)^2. \tag{51}$$

Considering the diagonal matrix elements, the above result implies

$$\langle jm\alpha|\Delta J_1^2|jm\alpha\rangle\langle jm\alpha|\Delta J_2^2|jm\alpha\rangle = \frac{|\alpha|^2}{4|\mathrm{Re}\alpha|^2}(\langle jm\alpha|J_3|jm\alpha\rangle)^2, \tag{52}$$

which will be $\frac{1}{4}(\langle jm\alpha|J_3|jm\alpha\rangle)^2$ *only when* α *is real*. Thus we have verified that from amongst the quasi-intelligent states, only the intelligent states satisfy equality in the Heisenberg uncertainty relation and for all other quasi-intelligent states

$$\Delta J_1^2\cdot\Delta J_2^2 > \tfrac{1}{4}(\langle J_3\rangle)^2$$

as expected.

Since in the above example, our interest was first to exemplify the efficiency of our method and second to reproduce equality in the Heisenberg uncertainty relation as a check on our methods, we have not tried to express our results in terms of the τ's though it could be done trivially.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

[1] C. Aragone, E. Chalbaud, and S. Salamo, J. Math. Phys. 17, 1963 (1976).

[2] See, for example, E. Merzbacher, in *Quantum Mechanics* (Wiley, New York, 1970), pp. 159–60.

[3] The normalization of the operators $J_3'(\alpha)$ and $J_\pm'(\alpha)$ have been chosen so as to satisfy Eqs. (15) and (16) which are exactly the same as satisfied by J_3, $J_\pm$. As one consequence of this normalization, we find that the spectrum of $J_3'(\alpha)$ for $\alpha \neq \pm 1$ coincides with the spectrum of J_3 and thus, for a given j, consists of the m's which satisfy $-j \leq m \leq j$ with unit steps.

[4] $\alpha = 0$ is a point which is simultaneously purely real and purely imaginary. Again $\alpha = 0 \Longrightarrow \tau \Longrightarrow 1 \Longrightarrow \theta = 2n\pi i$ and hence the corresponding states $|jm0\rangle$ are indeed obtainable from $|jm\rangle$ by a physical rotation. But this result is obvious since with $\alpha = 0$, $J_3'(0) = J_1$ which could obviously be obtained from J_3 by rotation. This special case has no features different from those of the Wigner states and is ignored in the sequel.

[5] See, e.g., footnote 8 in Ref. 1.

[6] The usual representation for the σ matrices is given by

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

[7] H. Bateman in *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, p. 238, Eq. (3).

[8] $J_3'(\alpha) = \frac{J_1 - i\alpha J_2}{(1-\alpha^2)^{1/2}} \Longrightarrow [J_3'(\alpha)]^\dagger = \frac{J_1 + i\alpha^* J_2}{[(1-\alpha^2)^{1/2}]^*}$.

The intelligent states. II. The computation of the Clebsch–Gordan coefficients

M. A. Rashid[a)]

International Centre for Theoretical Physics, Trieste, Italy
(Received 8 August 1977)

In this second of a series of papers on quasi-intelligent states, we give a general method for the computation of the Clebsch–Gordan coefficients for these states. In a special case, these coefficients are found to be closely related to the Clebsch–Gordan coefficients of the rotation group. We also discuss the nonuniqueness resulting naturally from the overcompleteness of these states.

1. INTRODUCTION

In the first paper of this series[1] (this paper will henceforth be referred to as I) we introduced the group-theoretic formulation for the study of the quasi-intelligent states which are generalizations of the states (called the intelligent states) satisfying equality in the Heisenberg uncertainty relation $\Delta J_1^2 \Delta J_2^2 \geq \frac{1}{4}|\langle J_3\rangle|^2$. In particular, we presented a method based on the knowledge of a certain generating function for the computation of matrix elements of polynomials in the infinitesimal generators of rotations in three dimensions between quasi-intelligent states.

In this paper, we continue this study and compute the Clebsch–Gordan coefficients for these states. Not surprisingly, these come out to be very closely related to the Clebsch–Gordan coefficients of the rotation group.

The present paper is organized as follows.

In Sec. 2, after redefining, for completeness, the various operators we compute their effect on the quasi-intelligent states. We utilize the results of this section in the next section to show that any Wigner state can be expressed as a linear combination of the quasi-intelligent states for any given complex number $\alpha \neq \pm 1$.[2] This is effectively the inversion of the expression for a quasi-intelligent state in terms of the Wigner states which was given in I. Indeed in Appendix A, we verify the correctness of this inversion. In Sec. 4, we derive the Clebsch–Gordan coefficients for the quasi-intelligent states and show that, up to a normalization independent of the magnetic quantum numbers, these Clebsch–Gordan coefficients for the same α are very closely related to the Clebsch–Gordan (CG) coefficients of the rotation group. In Appendix B, we show that this, so far, unknown normalization coefficient is indeed 1.

We emphasize that the quasi-intelligent states are eigenstates of a non-Hermitian operator having the same finite spectrum as the operator J_3 for a given angular momentum j. The non-Hermiticity of this operator makes the quasi-intelligent states non-orthogonal, thus some steps in the computation of the CG coefficients have to be handled rather carefully.

In Appendix C, we exemplify a consequence of the nonorthogonality of the quasi-intelligent states by showing that we *only* obtain a generalization of the expansion of the unit operator commonly known as completeness. Thus the quasi-intelligent states are, perhaps, not complete in this sense, though they are definitely complete in the sense that any Wigner state can be expressed as a linear combination of them.

2. THE OPERATORS $J_3'(\alpha)$, $J_\pm'(\alpha)$ AND THEIR EFFECT ON A QUASI-INTELLIGENT STATE

As in I, the *normalized* quasi-intelligent states corresponding to a given angular momentum j and a complex number $\alpha \neq \pm 1$ are given by[3]

$$|jm\alpha\rangle = [a^j{}_m(\alpha)]^{-1} \exp(\theta J_3) \exp(-i\tfrac{1}{2}\pi J_2)|jm\rangle, \tag{1}$$

where

$$e^\theta = \left(\frac{1-\alpha}{1+\alpha}\right)^{1/2} \tag{2}$$

and

$$a^j{}_m(\alpha) = \{\langle jm| \exp[-(\theta+\theta^*)J_1]|jm\rangle\}^{1/2}. \tag{3}$$

These states are eigenstates of the operator $J_3'(\alpha) = (J_1 - i\alpha J_2)/(1-\alpha^2)^{1/2}$. Indeed

$$J_3'(\alpha)|jm\alpha\rangle = m|jm\alpha\rangle. \tag{4}$$

Thus for a given j and a complex number $\alpha \neq \pm 1$, $J_3'(\alpha)$ has the spectrum $-j \leq m \leq j$, exactly the same as of the operator J_3.[4]

The basic difficulty in handling the quasi-intelligent states is the result of the obvious non-Hermiticity of the operator $J_3'(\alpha)$. As an immediate consequence, the corresponding eigenstates $|jm\alpha\rangle$ *might* not be orthogonal. Indeed it was explicitly verified in I that these states are not orthgonal for real $\theta \neq 0$. We also showed in I that for real $\theta \neq 0$, the normalization coefficients $a^j{}_m(\alpha) \neq 1$.

We define the operators

$$J_\pm'(\alpha) = \mp \frac{\alpha}{(1-\alpha^2)^{1/2}} J_1 \pm \frac{i}{(1-\alpha^2)^{1/2}} J_2 - J_3, \tag{5}$$

which together with $J_3'(\alpha)$ defined above satisfy the commutation relations

$$[J_3'(\alpha), J_\pm'(\alpha)] = \pm J_\pm'(\alpha), \tag{6}$$

$$[J_+'(\alpha), J_-'(\alpha)] = 2J_3'(\alpha), \tag{7}$$

which are exactly the same as those satisfied by J_3, $J_\pm = J_1 \pm iJ_2$.

Also

[a)] On leave of absence from Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria.

 0022-2488/78/1906-1397$1.00

$$\mathbf{J}^2 = J_1^2 + J_2^2 + J_3^2 = \tfrac{1}{2}[J_+J_- + J_-J_+] + J_3^2$$
$$= \tfrac{1}{2}[J'_+(\alpha)J'_-(\alpha) + J'_-(\alpha)J'_+(\alpha)] + J_3'^2(\alpha). \quad (8)$$

Now we wish to compute the effect of $J'_\pm(\alpha)$ on $|jm\alpha\rangle$. This can be immediately known from

$$\exp(-\theta J_3)J'_\pm(\alpha)\exp(\theta J_3) = \pm iJ_2 - J_3 \quad (9a)$$

and

$$\exp(i\tfrac{1}{2}\pi J_2)(\pm iJ_2 - J_3)\exp(-i\tfrac{1}{2}\pi J_2) = J_1 \pm iJ_2 = J_{\pm}. \quad (9b)$$

Thus,[5]

$$J'_\pm(\alpha)|jm\alpha\rangle = [a^j{}_m(\alpha)]^{-1}J'_\pm(\alpha)\exp(\theta J_3)\exp(-i\tfrac{1}{2}\pi J_2)|jm\rangle$$
$$= [a^j{}_m(\alpha)]^{-1}\exp(\theta J_3)\exp(-i\tfrac{1}{2}\pi J_2)J_\pm|jm\rangle$$
$$= \frac{a^j{}_{m\pm1}(\alpha)}{a^j{}_m(\alpha)}[(j\mp m)(j\pm m+1)]^{1/2}|j(m\pm1)\rangle. \quad (10)$$

We can immediately verify that Eqs. (4) and (10) are consistent with Eqs. (6)—(8).

In the following, we shall also require use of the states $|jm\alpha^c\rangle$. Here α^c is defined such that $\alpha \to -\alpha^*$, whereas $(1-\alpha^2)^{1/2} \to [(1-\alpha^2)^{1/2}]^*$ or in the language of θ, $\theta \to -\theta^*$. Using the variable τ, this operation[6] is expressed as $\tau \to (\tau^*)^{-1}$. The state $|jm\alpha\rangle$ is orthogonal to $|j'm'\alpha^c\rangle$. Indeed,

$$\langle j'm'\alpha^c|jm\alpha\rangle = \delta_{jj'}\delta_{mm'}[a^j{}_m(\alpha)]^{-2} \quad (11a)$$

since

$$a^j{}_m(\alpha) = a^j{}_m(\alpha^c), \quad (11b)$$

as can be deduced from Eq. (3) above or one may see it manifestly in Eq. (30b) in I.

3. EXPRESSION FOR A WIGNER STATE IN TERMS OF THE QUASI-INTELLIGENT STATES

In I, we derived the manifest expressions[7]

$$|jm\alpha\rangle = [a^j{}_m(\alpha)]^{-1}\left(\frac{(j-m)!}{(j+m)!}\right)^{1/2}\sum_{m_1,r}|jm_1\rangle\exp(m_1\theta)$$
$$\times\left(\frac{(j+m_1)!}{(j-m_1)!}\right)^{1/2}2^{-j+r}(-1)^r$$
$$\times\frac{(2j-r)!}{r!(j-m-r)!(j+m_1-r)!} \quad (12a)$$

$$= [a^j{}_m(\alpha)]^{-1}\left(\frac{(j+m)!}{(j-m)!}\right)^{1/2}\sum_{m_1,r}|jm_1\rangle\exp(m_1\theta)$$
$$\times\left(\frac{(j-m_1)!}{(j+m_1)!}\right)^{1/2}2^{-j+r}(-1)^{m_1-m+r}$$
$$\times\frac{(2j-r)!}{r!(j+m-r)!(j-m_1-r)!}, \quad (12b)$$

which are equivalent to the concise Eq. (1) in terms of the operation of the infinitesimal generators of the rotation group. Our purpose, in this section, is to utilize the results of the previous section to invert this equation to obtain an expression for any Wigner state as a linear combination of the quasi-intelligent states for any given complex $\alpha \neq \pm 1$.

To achieve it, let us go back to Eq. (1) which results in

$$|jm\rangle = a^j{}_m(\alpha)\exp(i\tfrac{1}{2}\pi J_2)\exp(-\theta J_3)|jm\alpha\rangle, \quad (13)$$

which is a *nonmanifest* form of the inversion we are seeking. Since we do know how $J'_3(\alpha)$, $J'_\pm(\alpha)$ operate on the quasi-intelligent states $|jm\alpha\rangle$, we attempt to express the operator $\exp(i\frac{1}{2}\pi J_2)\exp(-\theta J_3)$ in the form $\exp[aJ'_-(\alpha)]\exp[bJ'_3(\alpha)]\exp[cJ'_+(\alpha)]$. This can be done using the 2×2 representation $\sigma_i/2$ of the operators J_i. We find

$$a = -1, \quad e^b = \frac{1}{2\tau}, \quad c = \tau,$$

or

$$\exp\left(i\frac{\pi}{2}J_2\right)\exp(-\theta J_3) = \exp[-J'_-(\alpha)]\left(\frac{1}{2\tau}\right)^{J'_3(\alpha)}$$
$$\times\exp[\tau J'_+(\alpha)]. \quad (14)$$

Thus

$$|jm\rangle = a^j{}_m(\alpha)\exp[-J'_-(\alpha)]\left(\frac{1}{2\tau}\right)^{J'_3(\alpha)}\exp[\tau J'_+(\alpha)]|jm\alpha\rangle, \quad (15)$$

where for the operators on the right, we have already understood their operation on $|jm\alpha\rangle$ in the previous section. Utilizing Eqs. (4) and (10), we arrive at

$$|jm\rangle = \exp(-m\theta)\left(\frac{(j-m)!}{(j+m)!}\right)^{1/2}\sum_{m'm''}|jm''\alpha\rangle\left(\frac{(j-m'')!}{(j+m'')!}\right)^{1/2}$$
$$\times a^j{}_{m''}(\alpha)(-1)^{m'-m''}2^{-m'}\frac{(j+m')!}{(j-m')!(m'-m)!(m'-m'')!}. \quad (16)$$

In Appendix A, we shall explicitly verify the correctness of the above inversion.

In Eq. (16), we have expressed a given Wigner state $|jm\rangle$ as a linear combination of the quasi-intelligent states $|jm\alpha\rangle$ for a given complex $\alpha \neq \pm 1$ and $-j \leq m \leq j$. This shows that the quasi-intelligent states are indeed complete in the sense that any Wigner state can be expressed as a linear combination of them.

4. THE CLEBSCH-GORDAN COEFFICIENTS FOR THE QUASI-INTELLIGENT STATES

Now we have all the machinery at our disposal to enable us to compute the Clebsch—Gordan coefficients for the quasi-intelligent states. These coefficients are defined through the equation

$$|j_1m_1\alpha_1\rangle|j_2m_2\alpha_2\rangle$$
$$= \sum_{jm}(jm\alpha|j_1m_1\alpha_1;j_2m_2\alpha_2)|jm\alpha\rangle, \quad (17a)$$

where we have used *round brackets* to distinguish them from the usual Clebsch—Gordan coefficients of the rotation group. Equation (17) above expresses the completeness of the states $|jm\alpha\rangle$ for any complex $\alpha \neq \pm 1$ in the sense that any Wigner state $|jm\rangle$ can be expressed as a linear combination of them. Note particularly that, *in general*, the sum in Eq. (17) above is over both j and m where $|j_1 - j_2| \leq j \leq j_1 + j_2$ and $-j \leq m \leq j$ for a given j.

The Clebsch—Gordan coefficient $(jm\alpha|j_1m_1\alpha;j_2m_2\alpha_2)$ can be expressed as an inner product using the states $|jm\alpha^c\rangle$. Indeed

$$(jm\alpha\,|j_1m_1\alpha_1;j_2m_2\alpha_2)$$
$$=\langle jm\alpha^c|\,(|j_1m_1\alpha_1\rangle\,|j_2m_2\alpha_2\rangle)\rangle[a^j{}_m(\alpha)]^2 \tag{17b}$$

on making use of Eq. (11a). The inner product on the right above can be computed using Eqs. (12) and hence we can obtain a *general* Clebsch–Gordan coefficient for different values of α_1, α_2, and α. This coefficient is, however, very complicated and *cannot* be simplified except in special cases. In the following, we discuss a rather special case where $\alpha_1=\alpha_2=\alpha$. Now the definition of α^c has been chosen such that from

$$J_3'(\alpha)\,|jm\alpha\rangle = m\,|jm\alpha\rangle$$

one concludes

$$\langle jm\alpha^c|J_3'(\alpha)=m\langle jm\alpha^c|.$$

Thus Eq. (17b) implies, for the special case when $\alpha=\alpha_1=\alpha_2$,

$$(jm\alpha\,|j_1m_1\alpha_1;j_2m_2\alpha_2)$$
$$=\delta_{m,m_1+m_2}(jm\alpha\,|j_1m_1\alpha_1;j_2m_2\alpha_2), \tag{18}$$

or the summation in Eq. (17a) over m can be omitted with the understanding that $m=m_1+m_2$, i.e.,

$$|j_1m_1\alpha_1\rangle\,|j_2m_2\alpha_2\rangle$$
$$=\sum_j\,(j(m_1+m_2)\alpha\,|j_1m_1\alpha_1;j_2m_2\alpha_2)\,|jm\alpha\rangle. \tag{19}$$

Note that since the α's are kept the same throughout

$$J_3'^{(1)}(\alpha)+J_3'^{(2)}(\alpha)=\frac{J_1^{(1)}-i\alpha J_2^{(1)}}{(1-\alpha^2)^{1/2}}+\frac{J_1^{(2)}-i\alpha J_2^{(2)}}{(1-\alpha^2)^{1/2}}$$
$$=\frac{(J_1^{(1)}+J_1^{(2)})-i\alpha(J_2^{(1)}+J_2^{(2)})}{(1-\alpha^2)^{1/2}}$$
$$=\frac{J_1-i\alpha J_2}{(1-\alpha^2)^{1/2}}=J_3'(\alpha),$$

which results in the simplification given in Eq. (18) above and the ones which follow. Similar results hold for the operators $J_\pm'(\alpha)$.

Next we operate both sides of Eq. (19) by $J_+'(\alpha)=J_+'^{(1)}(\alpha)+J_+'^{(2)}(\alpha)$. Using Eq. (10), this operation gives

$$\frac{a^{j_1}{}_{(m_1+1)}(\alpha)}{a^{j_1}{}_{m_1}(\alpha)}[(j_1-m_1)(j_1+m_1+1)]^{1/2}\,|j_1(m_1+1)\alpha\rangle\,|j_2m_2\alpha\rangle$$
$$+\frac{a^{j_2}{}_{(m_2+1)}(\alpha)}{a^{j_2}{}_{m_2}(\alpha)}[(j_2-m_2)(j_2+m_2+1)]^{1/2}$$
$$\times|j_1m_1\alpha\rangle\,|j_2(m_2+1)\alpha\rangle$$
$$=\sum_j\frac{a^j{}_{(m+1)}(\alpha)}{a^j{}_m(\alpha)}[(j-m)(j+m+1)]^{1/2}$$
$$\times(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha)\,|j(m+1)\alpha\rangle.$$

Next we use Eq. (17) again for $|j_1(m_1+1)\alpha\rangle\,|j_2m_2\alpha\rangle$ and $|j_1m_1\alpha\rangle\,|j_2(m_2+1)\alpha\rangle$ which appear on the left-hand side of the above equation. This results in

$$\frac{a^{j_1}{}_{(m_1+1)}(\alpha)}{a^{j_1}{}_{m_1}(\alpha)}[(j_1-m_1)(j_1+m_1+1)]^{1/2}$$
$$\times\sum_j\,(j(m+1)\alpha\,|j_1(m_1+1)\alpha;j_2m_2\alpha)\,|j(m+1)\alpha\rangle$$
$$+\frac{a^{j_2}{}_{(m_2+1)}(\alpha)}{a^{j_2}{}_{m_2}(\alpha)}[(j_2-m_2)(j_2+m_2+1)]^{1/2}$$
$$\times\sum_j\,(j(m+1)\alpha\,|j_1m_1\alpha;j_2(m_2+1)\alpha)\,|j(m+1)\alpha\rangle$$
$$=\sum_j\frac{a^j{}_{(m+1)}(\alpha)}{a^j{}_m(\alpha)}[(j-m)(j+m+1)]^{1/2}$$
$$\times(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha)\,|j(m+1)\alpha\rangle. \tag{20}$$

In order to remove the sum over j we take the inner product of both sides with $\langle j'(m+1)\alpha|$ and use $\langle j'(m+1)\alpha\,|j(m+1)\alpha\rangle=\delta_{jj'}$. [Note that this orthonormality involves j and not m and hence does not depend upon the non-Hermiticity of $J_3'(\alpha)$.] We also multiply by $a^{j_1}{}_{m_1}(\alpha)a^{j_2}{}_{m_2}(\alpha)[a^j{}_{(m+1)}(\alpha)]^{-1}$ to arrive at

$$\frac{a^{j_1}{}_{(m_1+1)}(\alpha)a^{j_2}{}_{m_2}(\alpha)}{a^j{}_{(m+1)}(\alpha)}[(j_1-m_1)(j_1+m_1+1)]^{1/2}$$
$$\times(j(m+1)\alpha\,|j_1(m_1+1)\alpha;j_2m_2\alpha)$$
$$+\frac{a^{j_1}{}_{m_1}(\alpha)a^{j_2}{}_{(m_2+1)}(\alpha)}{a^j{}_{m+1}(\alpha)}[(j_2-m_2)(j_2+m_2+1)]^{1/2}$$
$$\times(j(m+1)\alpha\,|j_1m_1\alpha;j_2(m_2+1)\alpha)$$
$$=\frac{a^{j_1}{}_{m_1}(\alpha)a^{j_2}{}_{m_2}(\alpha)}{a^j{}_m(\alpha)}[(j-m)(j+m+1)]^{1/2}$$
$$\times(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha), \tag{21}$$

which shows that the quantities

$$\frac{a^{j_1}{}_{m_1}(\alpha)a^{j_2}{}_{m_2}(\alpha)}{a^j{}_m(\alpha)}(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha)$$

satisfy the same recursion relation as the one satisfied by the CG coefficients $\langle jm\,|j_1m_1;j_2m_2\rangle$ of the rotation group. Hence we conclude that

$$(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha)$$
$$=\beta(j_1j_2j;\alpha)\frac{a^j{}_m(\alpha)}{a^{j_1}{}_{m_1}(\alpha)a^{j_2}{}_{m_2}(\alpha)}\langle jm\,|j_1m_1;j_2m_2\rangle, \tag{22}$$

where the coefficient $\beta(j_1j_2j;\alpha)$ will have to be fixed by normalization and choice of phase.[8] These coefficients are independent of the magnetic quantum numbers. Indeed, in Eq. (22) we have been able to separate the dependence of the CG coefficient for the intelligent states into the corresponding CG coefficient of the rotation group and the normalization factors of the involved intelligent states.

To calculate the β's, it is clear from the above equation that if we could obtain the coefficient on the left for some special values of the magnetic quantum numbers, we would be able to obtain the β in this equation. Note that Eq. (19) implies that the CG coefficient $(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha)$ can be obtained by taking the inner product of $|j_1m_1\alpha\rangle\,|j_2m_2\alpha\rangle$ with $\langle jm\alpha|$. In Appendix B we shall carry out this program and show that the β's can, in fact, be chosen to be just one. Thus we find finally

$$(jm\alpha\,|j_1m_1\alpha;j_2m_2\alpha)$$
$$=\frac{a^j{}_m(\alpha)}{a^{j_1}{}_{m_1}(\alpha)a^{j_2}{}_{m_2}(\alpha)}\langle jm\,|j_1m_1;j_2m_2\rangle, \tag{23}$$

where both sides are identically zero if $m\neq m_1+m_2$.

At this stage, we wish to remark on the possible nonuniqueness in the expansion for the product

$|j_1 m_1 \alpha_1\rangle |j_2 m_2 \alpha_2\rangle$ in terms of the states $|jm\alpha\rangle$. Already in Eq. (17a) wherein we defined the Clebsch–Gordan coefficients, we have the built-in nonuniqueness since the complex "α" appearing on the right is at our disposal. Note that the states $|jm\alpha\rangle$ are complete for each $\alpha \neq 1$. Considering the set of all "α" at our disposal, we possess a highly *overcomplete* set of vectors which should naturally result in the nonuniqueness expressed above. In the special case expressed in Eq. (19), we have restricted ourselves to $\alpha = \alpha_1$ (note that $\alpha_1 = \alpha_2$). Using the states $|jm\alpha'\rangle$ and Eq. (III. 6) in Appendix C we can rewrite Eq. (19) as

$$|j_1 m_1 \alpha\rangle |j_2 m_2 \alpha\rangle = \sum_{jm} (j(m_1+m_2)\alpha\,|j_1 m_1 \alpha; j_2 m_2 \alpha) \times \langle jm\alpha'^c | j(m_1+m_2)\alpha\rangle [a^j{}_m(\alpha')]^2 |jm\alpha'\rangle, \tag{24}$$

which is an expansion as a linear combination of $|jm\alpha'\rangle$ and reduces to Eq. (19) in case $\alpha = \alpha'$ on using Eq. (11a). Note that the expression on the right in the above equation has the additional (perhaps artificial in this special case) summation over m.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX A: EXPLICIT VERIFICATION OF THE INVERSION IN EQ. (16)

We proved Eq. (16) in Sec. 3 by using the knowledge of the effect of the operators $J_\pm'(\alpha)$ on $|jm\alpha\rangle$. Now we verify that Eq. (16) indeed provides an inversion to Eq. (12). For this purpose, we substitute for $|jm''\alpha\rangle$ in Eq. (16) from Eq. (12). This makes the expression E on the right-hand side of Eq. (16) become

$$E = \exp(-m\theta)\left[\frac{(j-m)!}{(j+m)!}\right]^{1/2} \sum_{m_1 m_2 m_3 r} |jm_3\rangle (-1)^{m_1-2m_2+m_3+r} \times \exp(m_3\theta) 2^{-j-m_1+r} \left[\frac{(j-m_3)!}{(j+m_3)!}\right]^{1/2} \times \frac{(j+m_1)!\,(2j-r)!}{(j-m_1)!\,(m_1-m)!\,(m_1-m_2)!\,r!\,(j+m_2-r)!\,(j-m_3-r)!}. \tag{A1}$$

Though m_1, m_2, m_3 summations may not be over integers, before performing any summation, we shall ensure that the variable we choose is indeed an integer. Now the m_2 summation can be performed [by first replacing say $(m_1 - m_2)$ by a new variable in place of m_2 and summing over it]. This gives

$$E = \exp(-m\theta)\left[\frac{(j-m)!}{(j+m)!}\right]^{1/2} \sum_{m_1 m_3 r} |jm_3\rangle (-1)^{m_3-m_1+r} \times \exp(m_3\theta)\left[\frac{(j-m_3)!}{(j+m_3)!}\right]^{1/2} \times \frac{(j+m_1)!\,(2j-r)!}{(j-m_1)!\,(m_1-m)!\,r!\,(j+m_1-r)!\,(j-m_3-r)!}. \tag{A2}$$

The r summation can now be done which results in

$$E = \exp(-m\theta)\left[\frac{(j-m)!}{(j+m)!}\right]^{1/2} \sum_{m_1 m_3} |jm_3\rangle (-1)^{m_3-m_1} \exp(m_3\theta) \times \left[\frac{(j+m_3)!}{(j-m_3)!}\right]^{1/2} \frac{1}{(m_1-m)!\,(-m_1+m_3)!}. \tag{A3}$$

Now the m_1 summation produces δ_{mm_3} which finally gives

$$E = |jm\rangle,$$

exactly the same as on the left-hand side of Eq. (16).

APPENDIX B: EVALUATION OF THE NORMALIZATION COEFFICIENT $\beta(j_1 j_2 j;\alpha)$

In this Appendix we wish to compute the β's that appeared in Eq. (22) relating the CG coefficients for the quasi-intelligent states with those of the rotation group. As remarked earlier, we should be able to compute β's from the knowledge of the CG coefficient on the left for some special choice of values of the magnetic quantum numbers. From the representations for $|jm\alpha\rangle$ given in Eq. (12) we can, indeed, give an explicit answer for these CG coefficients in terms of the CG coefficients of the rotation group. In general, however, manifesting the factorization expressed in Eq. (22) must be a formidable task. For a very special case, we might, hopefully, be lucky. This is fortunately the case for the special choice $m_1 = j_1$, $m_2 = -j_2$, $m = j_1 - j_2$. Note that since $|j_1 - j_2| \leq j \leq j_1 + j_2$ the value $j_1 - j_2$ is indeed a permissible value. From the representations in Eq. (12) we obtain

$$|j_1 j_1 \alpha\rangle = [a^{j_1}{}_{j_1}(\alpha)]^{-1} 2^{-j_1} [(2j_1)!\,]^{1/2} \sum_{m_1'} |j_1 m_1'\rangle \frac{\exp(m_1'\theta)}{[(j_1+m_1')!\,(j_1-m_1')!\,]^{1/2}}, \tag{B1}$$

$$|j_2(-j_2)\alpha\rangle = [a^{j_2}{}_{(-j_2)}(\alpha)]^{-1} 2^{-j_2} [(2j_2)!\,]^{1/2} \sum_{m_2'} |j_2 m_2'\rangle (-1)^{j_2+m_2'} \frac{\exp(m_2'\theta)}{[(j_2+m_2')!\,(j_2-m_2')!\,]^{1/2}}, \tag{B2}$$

and

$$\langle j(j_1-j_2)\alpha| = [a^j{}_{(j_1-j_2)}(\alpha)]^{-1} 2^{-j} \left[\frac{(j-j_1+j_2)!}{(j+j_1-j_2)!}\right]^{1/2} \sum_{m'r} \langle jm'| \left[\frac{(j+m')!}{(j-m')!}\right]^{1/2} (-2)^r \frac{\exp(m'\theta^*)(2j-r)!}{r!\,(j-j_1+j_2-r)!\,(j+m'-r)!}. \tag{B3}$$

Thus

$$(j(j_1-j_2)\alpha\,|j_1 j_1 \alpha; j_2(-j_2)\alpha)$$

$$= [a^{j_1}{}_{j_1}(\alpha) a^{j_2}{}_{(-j_2)}(\alpha) a^{j}{}_{(j_1-j_2)}(\alpha)]^{-1} 2^{-(j_1+j_2+j)} \left[\frac{(2j_1)!\,(2j_2)!\,(j-j_1+j_2)!}{(j+j_1-j_2)!}\right]^{1/2} \sum_{m_1' m_2' r} \langle jm' | j_1(m'-m_2'); j_2 m_2' \rangle$$

$$\times \left[\frac{(j+m')!}{(j_1+m'-m_2')!\,(j_1-m'+m_2')!\,(j_2+m_2')!\,(j_2-m_2')!\,(j-m')!}\right]^{1/2} 2^r (-1)^{j_2+m_2'+r} \exp[(m'(\theta+\theta^*)]$$

$$\times \frac{(2j-r)!}{r!\,(j-j_1+j_2-r)!\,(j+m'-r)!}\,, \tag{B4}$$

where knowing that the CG coefficient $\langle jm' | j_1 m_1'; j_2 m_2' \rangle$ can be nonzero only when $m' = m_1' + m_2'$, we have eliminated the summation over m_1' by replacing m_1' by $m' - m_2'$ everywhere. Now we use[9]

$$\langle jm' | j_1(m'-m_2'); j_2 m_2' \rangle$$

$$= \left[\frac{(2j+1)(j-j_1+j_2)!\,(j_1+m'-m_2')!\,(j_1-m'+m_2')!\,(j_2-m_2')!\,(j+m')!}{(j+j_1+j_2+1)!\,(-j+j_1+j_2)!\,(j+j_1-j_2)!\,(j_2+m_2')!\,(j-m')!}\right]^{1/2}$$

$$\times \sum_s (-1)^{-j+j_1+j_2+s} \frac{(j_2+m_2'+s)!\,(j+j_1-m_2'-s)!}{s!\,(j_1+m'-m_2'-s)!\,(j_2-m_2'-s)!\,(j-j_1+m_2'+s)!} \tag{B5}$$

for the CG coefficient of the rotation group which appears above in Eq. (B4). This results is

$$(j(j_1-j_2)\alpha\,|j_1 j_1 \alpha; j_2(-j_2)\alpha)$$

$$= [a^{j_1}{}_{j_1}(\alpha) a^{j_2}{}_{(-j_2)}(\alpha) a^{j}{}_{(j_1-j_2)}(\alpha)]^{-1} \left[\frac{(2j+1)(2j_1)!\,(2j_2)!}{(j+j_1+j_2+1)!\,(j_1+j_2-j)!}\right]^{1/2} \left\{\frac{(j-j_1+j_2)!}{(j+j_1-j_2)} \sum_{m' m_2' r s} (-1)^{-j+j_1+2j_2+m_2'+r+s} 2^{-(j+j_1+j_2)+r}\right.$$

$$\left. \times \frac{(j+m')!}{(j-m')!\,(j_2+m_2')!} \exp[m'(\theta+\theta^*)] \frac{(2j-r)!\,(j_2+m_2'+s)!\,(j+j_1-m_2'-s)!}{r!\,(j-j_1+j_2-r)!\,(j+m'-r)!\,s!\,(j_1+m'-m_2'-s)!\,(j_2-m_2'-s)!\,(j-j_1+m_2'+s)!}\right\}. \tag{B6}$$

The expression within the curly brackets is called S in the following. To simplify S, we replace $j_2 - m_2' - s$ by s, i.e., we replace s by $j_2 - m_2' - s$. This gives

$$S = \frac{(j-j_1+j_2)!}{(j+j_1-j_2)!} \sum_{m' m_2' r s} 2^{-(j+j_1+j_2)+r} (-1)^{-j+j_1-j_2+r+s} \frac{(j+m')!}{(j_2+m_2')!\,(j-m')!} \exp[m'(\theta+\theta^*)]$$

$$\times \frac{(2j-r)!\,(2j_2-s)!\,(j+j_1-j_2+s)!}{r!\,(j-j_1+j_2-r)!\,(j+m'-r)!\,s!\,(j_1-j_2+m'+s)!\,(j_2-m_2'-s)!\,(j-j_1+j_2-s)!}\,. \tag{B7}$$

The m_2' summation now gives $2^{2j_2-s}/(2j_2-s)!$ This results in

$$S = \frac{(j-j_1+j_2)!}{(j+j_1-j_2)!} \sum_{m' r s} 2^{-(j+j_1-j_2)+r-s} (-1)^{-j+j_1-j_2+r+s} \frac{(j+m')!}{(j-m')!} \exp[m'(\theta+\theta^*)]$$

$$\times \frac{(2j-r)!\,(j+j_1-j_2+s)!}{r!\,(j-j_1+j_2-r)!\,(j+m'-r)!\,s!\,(j_1-j_2+m'+s)!\,(j-j_1+j_2-s)!}\,. \tag{B8}$$

To put S in a form which can be recognized, we replace $j - j_1 + j_2 - s$ by a new variable s which results in

$$S = \frac{(j-j_1+j_2)!}{(j+j_1-j_2)!} \sum_{m' r s} \frac{(j+m')!}{(j-m')!} \exp[m'(\theta+\theta^*)] 2^{-2j+r+s} (-1)^{r+s}$$

$$\times \frac{(2j-r)!\,(2j-s)!}{r!\,(j-j_1+j_2-r)!\,(j+m'-r)!\,s!\,(j-j_1+j_2-s)!\,(j+m'-s)!}\,. \tag{B9}$$

Comparing the above form of S with Eq. (12a), we immediately conclude

$$S = [a^{j}{}_{(j_1-j_2)}(\alpha)]^2.$$

Now we return to Eq. (B6). Recognizing that

$$\langle j(j_1-j_2)\,|j_1 j_1; j_2(-j_2)\rangle = \left(\frac{(2j+1)(2j_1)!\,(2j_2)!}{(j+j_1+j_2+1)!\,(j_1+j_2-j)!}\right)^{1/2}, \tag{B10}$$

we find

$$(j(j_1-j_2)\alpha\,|j_1 j_1 \alpha; j_2(-j_2)\alpha) = \frac{a^{j}{}_{(j_1-j_2)}(\alpha)}{a^{j_1}{}_{j_1}(\alpha) a^{j_2}{}_{(-j_2)}(\alpha)} \langle j(j_1-j_2)\,|j_1 j_1; j_2(-j_2)\rangle, \tag{B11}$$

which on comparison with Eq. (22) shows that $\beta(j_1 j_2 j; \alpha) = 1$. Note that it has been fixed completely by the *phase convention* used in *defining* the relationship [Eqs. (12)] between the quasi-intelligent states and the Wigner states. This equation has a built in *phase convention* which *cannot* be fixed by knowing *only* that $|jm\alpha\rangle$ is an eigenstate of the operator $J_3'(\alpha)$.

In the above computations, we have used the fact that the Clebsch–Gordan coefficient $(j(j_1-j_2)\alpha\,|j_1 j_1 \alpha; j_2(-j_2)\alpha)$

is just the inner product $\langle j(j_1-j_2)\alpha\,|\,(|j_1j_2\alpha\rangle\,|j_2(-j_2)\alpha\rangle)\rangle$. We could have also used that this coefficient is also equal to the inner product $\langle j(j_1-j_2)\alpha^c\,|\,(|j_1j_1\alpha\rangle\,|j_2(-j_2)\alpha\rangle)\rangle[a^j{}_{j_1-j_2}(\alpha)]^2$. It is obvious from (B9) above that the S corresponding to this new inner product would have been 1 and we would be able to reproduce the previous results.

APPENDIX C: AN ANALOG OF THE EXPANSION OF THE UNIT OPERATOR

In this Appendix, we shall prove that

$$\sum_m |jm\alpha\rangle\langle jm\alpha|\,[a^j{}_m(\alpha)]^2=\sum_m |jm\rangle\langle jm|\,\exp[m(\theta+\theta^*)]. \tag{C1}$$

Using Eq. (12a) for $|jm\alpha\rangle$ and Eq. (12b) for $\langle jm\alpha|$ we get

$$\sum_m |jm\alpha\rangle\langle jm\alpha|\,[a^j{}_m(\alpha)]^2=\sum_{mm_1m_2rs} |jm_1\rangle\langle jm_2|\,\exp(m_1\theta+m_2\theta^*)\left[\frac{(j+m_1)!\,(j-m_2)!}{(j-m_1)!\,(j+m_2)!}\right]^{1/2}2^{-2j+r+s}$$
$$\times(-1)^{m_2-m+r+s}\frac{(2j-r)!\,(2j-s)!}{r!\,(j-m-r)!\,(j+m_1-r)!\,s!\,(j+m-s)!\,(j-m_2-s)!}\,. \tag{C2}$$

Now

$$\sum_m (-1)^{j-m-r}\frac{1}{(j-m-r)!\,(j+m-s)!}=\delta_{2j-r-s,0}. \tag{C3}$$

From the above two equations, we obtain

$$\sum_m |jm\alpha\rangle\langle jm\alpha|\,[a^j{}_m(\alpha)]^2=\sum_{m_1m_2r} |jm_1\rangle\langle jm_2|\,\exp(m_1\theta+m_2\theta^*)\left[\frac{(j+m_1)!\,(j-m_2)!}{(j-m_1)!\,(j+m_2)!}\right]^{1/2}(-1)^{j+m_2-r}$$
$$\times\frac{1}{(j+m_1-r)!\,(-j-m_2+r)!}\,. \tag{C4}$$

On performing the trivial r summation, we obtain a delta function $\delta_{m_1m_2}$ and finally arrive at Eq. (C1) which is a *generalization* of the expansion of a unit operator in the sense that if we restricted ourselves to real $\theta=0$, we would obtain as a *special* case of Eq. (C1)

$$\sum_m |jm\alpha\rangle\langle jm\alpha|=\sum_m |jm\rangle\langle jm|=I, \tag{C5}$$

which is indeed an expansion of the unit operator. We did use the fact that for real $\theta=0$, $a^j{}_m(\alpha)=1$. But in this case, the states $|jm\alpha\rangle$ are indeed *orthonormalized* by just writing

$$|jm\alpha\rangle=\exp(\theta J_3)\exp(-i\tfrac{1}{2}\pi J_2)\,|jm\rangle$$

and Eq. (C5) should be obvious.

Incidentally, a proof similar to the above results in

$$\sum_m |jm\alpha^c\rangle\langle jm\alpha|\,[a^j{}_m(\alpha)]^2=\sum_m |jm\rangle\langle jm|=I. \tag{C6}$$

[1] M.A. Rashid, J. Math. Phys. **19**, 1391 (1978).
[2] This restriction is explained in Ref. 1.
[3] See Eq. (29) in Ref. 1.
[4] The normalization of the operator $J_3'(\alpha)$ has been chosen to have the same spectrum as that of J_3.
[5] Note that our normalization of the operators makes the answers of their effects on the quasi-intelligent states as very simple.
[6] Note that $J_3'(\alpha^c)=[J_3'(\alpha)]^\dagger$, whereas $J_\pm'(\alpha^c)=[J_\mp'(\alpha)]^\dagger$.
[7] See Eqs. (32b) and (32c) in Ref. 1.
[8] We shall make this choice by invoking consistency with Eqs. (12).
[9] $\langle jm'\,|\,j_1(m'-m_2');j_2m_2'\rangle$
$=(-1)^{j_1+j_2-j}(-1)^{j_2+m_2'}\left(\frac{2j+1}{2j_1+1}\right)\langle j_2m_2';j(-m')\,|\,j_1(-m'+m_2')\rangle$
on combining Eqs. (3.5.15) and (3.5.17) in A.R. Edmonds, in *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1957). Finally we used Eq. (3.6.10) for the CG coefficient on the right. This premanipulation of the CG coefficient has reduced the size of the Appendix considerably.

Three- and four-wave interactions in plasmas

T. J. M. Boyd and J. G. Turner[a)]

University of Wales, U.C.N.W., Bangor, Wales
(Received 31 May 1977)

The nonlinear interaction of three and four waves in a warm magnetized plasma is studied using the Lagrangian formalism. General expressions for the coupling coefficients are obtained in each case. Coupled-mode equations describing four-wave interactions have been derived and solved. The extension of the theory to the interaction of random phase waves and weak inhomogeneity is also discussed.

1. INTRODUCTION

The study of nonlinear plasma wave interactions is of fundamental importance in understanding the dynamics of a weakly turbulent plasma and since Sturrock[1] first investigated the interaction of three one-dimensional plasma oscillations, increasing interest has focused on this aspect of plasma research. Applications of these nonlinear processes include studies of nonthermal plasma radiation, plasma diagnostics, and parametric decay processes. The latter are of current interest in connection with laser fusion studies.[2]

The literature now abounds with studies of three-wave interactions where the usual approach adopted is the examination of *specific* triads of interacting waves using either conventional perturbation analysis of the governing equations of motion (whether a fluid or a statistical mechanical description) or, less commonly, the alternative based on the Lagrangian density for a plasma first formulated by Low.[3]

Many references to work adopting the former approach are to be found in the monograph by Davidson.[4] A particularly clear account of the role of the Lagrangian approach to problems in nonlinear wave theory has been given by Whitham[5] and its particular application to problems involving plasma waves may be found in the work of Boyd and Turner[6,7] and of Dougherty.[8,9]

Conservation of energy and momentum for these interactions demands that the frequencies ω_j $(\omega_j > 0)$, $j = 1, 2, 3$ and corresponding wave vectors $\mathbf{k}_j$ satisfy the synchronism conditions

$$\omega_1 + \omega_2 \simeq \omega_3, \tag{1}$$

$$\mathbf{k}_1 + \mathbf{k}_2 \simeq \mathbf{k}_3 \tag{2}$$

where $\omega_j(\mathbf{k}_j)$ is the solution of the appropriate linear dispersion relation. In the linear regime the solution of the governing equations of motion for the electric field $\mathbf{E}$ is given by the superposition of plane waves of the form $\mathbf{E} = \mathrm{Re}[\mathbf{E}_0 \exp i(\mathbf{k}\cdot\mathbf{x} - \omega t)]$ with $\mathbf{E}_0$ constant. An important assumption of the theory is that the fields described by the nonlinear equations can be approximately represented by $\mathbf{E}(\mathbf{x}, t) = \mathrm{Re}[\hat{\mathbf{E}}(\mathbf{x}, t)\exp i(\mathbf{k}\cdot\mathbf{x} - \omega t)]$ where $\hat{\mathbf{E}}(\mathbf{x}, t)$ is a slowly varying function of $\mathbf{x}$ and t, for which only the first derivatives with respect to $\mathbf{x}$ and t need be included. This description is equivalent to the Bogoliubov–Krylov[10] multiple scale expansion of the independent variables. In the well defined phase approximation (WDPA) corresponding to the criterion $\Delta\omega\tau \ll 1$ where $\Delta\omega = \omega_3 - \omega_2 - \omega_1$ is the frequency mismatch and τ is some characteristic time of the physical process, the analysis leads to a set of coupled-mode equations

$$\begin{aligned} i\omega_{1,2}^{-1}\left(\frac{\partial}{\partial t} + \mathbf{v}_{g1,2}\cdot\nabla\right)\hat{E}_{1,2} &= \frac{\Gamma_{wc}^*}{\Gamma_{1,2}}\hat{E}_{2,1}^*\hat{E}_3, \\ i\omega_3^{-1}\left(\frac{\partial}{\partial t} + \mathbf{v}_{g3}\cdot\nabla\right)\hat{E}_3 &= \frac{\Gamma_{wc}}{\Gamma_3}\hat{E}_1\hat{E}_2, \end{aligned} \tag{3}$$

where $\mathbf{v}_{gn} \equiv \partial\omega_n/\partial\mathbf{k}_n$ is the group velocity of wave n and $\Gamma_{wc}/\Gamma_{1,2,3}$ are coupling coefficients.

It is now logical to include all existing work on wave–wave interactions within a general theory of nonlinear wave coupling and in this paper we derive coupling coefficients for arbitrary three- and four-wave interactions in a warm magnetoactive plasma. In Sec. 2 a brief outline of the Lagrangian approach to nonlinear wave processes is presented and a general expression for the coupling coefficient Γ_{wc} describing three-wave interactions is given in Sec. 3. In Sec. 4, the corresponding results for four-wave interactions are derived. Coupled-mode equations are also obtained using quantum mechanical considerations, and solutions of the coupled-mode equations are found. The paper concludes with a discussion in which the results are applied to the interaction between four longitudinal waves in a warm isotropic plasma.

2. LAGRANGIAN THEORY

In this section a brief account of the Lagrangian approach to nonlinear plasma wave interactions is presented. A detailed account of the procedure is to be found in Boyd and Turner.[6] The Lagrangian, first formulated by Low,[3] for a warm plasma in a magnetic field $\mathbf{B}_0$ is

$$L = \int\int \mathcal{L}\, d\mathbf{x}\, d\mathbf{v},$$

where $\mathcal{L}$ is the Lagrangian density given by

$$\begin{aligned} \mathcal{L} = \sum_{\alpha=e,i} n_{0\alpha} f_{0\alpha}(\mathbf{x}, \mathbf{v})\left\{\frac{m_\alpha}{2}(D_{\Omega\alpha}\mathbf{r}_\alpha)^2 - q_\alpha\left(\phi - \frac{\mathbf{v}\circ\mathbf{A}}{c}\right)\right\} \\ + \frac{\chi(\mathbf{v})}{8\pi}\left[\left(\nabla\phi + \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}\right)^2 - (\nabla\wedge\mathbf{A})^2\right], \end{aligned} \tag{4}$$

where $\alpha = e, i$ denotes the electron, ion contributions, $n_{0\alpha}$ is the equilibrium number density, and $f_{0\alpha}$ is the equilibrium distribution function normalized to unity. m_α and q_α are the mass and charge of species α re-

[a)]Permanent address: Polytechnic of Central London, 115 New Cavendish Street, London W1M 8JS, England.

 0022-2488/78/1906-1403$1.00

spectively, $\mathbf{r}_\alpha$ is the particle displacement, and ϕ and $\mathbf{A}$ are the electrostatic, and magnetic vector potentials, respectively. The operator $D_{\Omega\alpha} = \partial/\partial t + \mathbf{v}\cdot\Delta + \mathbf{v}\wedge\Omega_\alpha \cdot \partial/\partial\mathbf{v}$, where $\Omega_\alpha = q_\alpha \mathbf{B}_0/m_\alpha c$ and $|\Omega_\alpha|$, is the particle cyclotron frequency. $\chi(\mathbf{v})$ is an arbitrary function of velocity such that $\int\chi(\mathbf{v})\,d\mathbf{v} = 1$. The electric and magnetic fields are related to the potentials by the relations

$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla\wedge\mathbf{A}.$$

The generalized variables for the Lagrangian density are $\mathbf{r}_\alpha$, ϕ, and $\mathbf{A}$. The associated energy density $\mathcal{H}$ is given by

$$\mathcal{H} = \partial\mathcal{L}\Big/\partial\left(\frac{\partial\mathbf{r}_\alpha}{\partial t}\right)\cdot\frac{\partial\mathbf{r}_\alpha}{\partial t} + \partial\mathcal{L}\Big/\partial\left(\frac{\partial\mathbf{A}}{\partial t}\right)\cdot\frac{\partial\mathbf{A}}{\partial t} - \mathcal{L}.$$

Following Low's procedure we write $\mathbf{r} = \mathbf{r}_0 + \epsilon\mathbf{r}^{(1)}$, $\phi = \epsilon\phi^{(1)}$, and $\mathbf{A} = \mathbf{A}_0 + \epsilon\mathbf{A}^{(1)}$, where the subscript 0 refers to the equilibrium state and the superscript (1) denotes the perturbation of this state due to the waves. A formal expansion of $\mathcal{L}$ and $\mathcal{H}$ may now be made in powers of the perturbation ϵ,

$$\mathcal{L} = \mathcal{L}_0 + \epsilon\mathcal{L}_1 + \epsilon^2\mathcal{L}_2 + \epsilon^3\mathcal{L}_3 + \epsilon^4\mathcal{L}_4 + \cdots$$

with a corresponding expansion for $\mathcal{H}$. $\mathcal{L}_0$ involves equilibrium quantities only and $\mathcal{L}_1$ has no effect on the equations of motion for the first-order quantities. The variation of $\mathcal{L}_2$ with respect to the generalized variables yields the linear Vlasov–Maxwell equations describing plasma wave propagation. Wave–wave interactions are then described by third and higher order terms. For three-wave interactions, the perturbation series is truncated after $\mathcal{L}_3$ whilst the inclusion of $\mathcal{L}_4$ allows a description of four-wave interaction processes. Dropping the superscript (1) for convenience, the expressions for $\mathcal{L}_3$ and $\mathcal{L}_4$ are given by

$$\mathcal{L}_3 = \sum_\alpha n_{0\alpha} f_{0\alpha}\left(-\tfrac{1}{2}q_\alpha(\mathbf{r}_\alpha\cdot\nabla)^2\phi + \frac{q_\alpha}{2c}\mathbf{v}\cdot(\mathbf{r}_\alpha\cdot\nabla)^2\mathbf{A} + \frac{q_\alpha}{c}D_{\Omega\alpha}\mathbf{r}_\alpha\cdot(\mathbf{r}_\alpha\cdot\nabla)\mathbf{A}\right), \tag{5}$$

$$\mathcal{L}_4 = \sum_\alpha n_{0\alpha} f_{0\alpha}\left(-\tfrac{1}{6}q_\alpha(\mathbf{r}_\alpha\cdot\nabla)^3\phi + \frac{q_\alpha}{6c}\mathbf{v}\cdot(\mathbf{r}_\alpha\cdot\nabla)^3\mathbf{A} + \frac{q_\alpha}{2c}D_{\Omega\alpha}\mathbf{r}_\alpha\cdot(\mathbf{r}_\alpha\cdot\nabla)^2\mathbf{A}\right). \tag{6}$$

In the case of a linear wave, for some parameter of wave n

$$U_n = \mathrm{Re}[\hat{U}_n \exp i(\mathbf{k}_n\cdot\mathbf{x} - \omega_n t)], \tag{7}$$

with $\hat{U}_n$ constant in space and time. To describe wave coupling it is now assumed that solutions of the nonlinear equations have the form (7) but with $\hat{U}_n$ now being regarded as a slowly varying function of $\mathbf{x}$ and t, so that

$$U_n = \mathrm{Re}[\hat{U}_n(\mathbf{x}, t)\exp i(\mathbf{k}_n\cdot\mathbf{x} - \omega_n t)]. \tag{8}$$

The perturbations are now separated into their individual wave components, so that for general three-, four-wave interactions,

$$\mathbf{r}_\alpha = \sum_{n=1}^{3;4}\mathbf{r}_{n\alpha}, \quad \phi = \sum_{n=1}^{3;4}\phi_n, \quad \mathbf{A} = \sum_{n=1}^{3;4}\mathbf{A}_n, \tag{9}$$

with each $\mathbf{r}_{n\alpha}$, ϕ_n, $\mathbf{A}_n$ given by expressions of the form (8). We now introduce a space–time averaging procedure (denoted by a bar) taken over intervals of space and periods of time which are long compared with the periods of oscillation k_n^{-1}, ω_n^{-1}, but short compared with intervals and periods over which the amplitudes $\hat{U}_n$ vary appreciably. Choosing gauge potentials such that the scalar potential in the transverse waves vanishes, then

$$\mathbf{E}^L = -\nabla\phi, \quad \mathbf{E}^T = -\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla\wedge\mathbf{A}.$$

Finally from $\mathcal{L}_2$ we derive the linear equations of motion which enable us to express the $\{\mathbf{r}_{n\alpha}, \phi_n, \mathbf{A}_n\}$ in terms of one wave parameter. In our case, we choose this to be the electric field wave amplitude of each wave E_n. The variation of $\mathcal{L}_2$ with respect to $\mathbf{r}_\alpha$ yields the first-order Lorentz equation

$$m_\alpha D_{\Omega\alpha}^2\mathbf{r}_\alpha = q_\alpha\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\wedge\mathbf{B}\right) + \frac{q_\alpha}{c}D_{\Omega\alpha}\mathbf{r}_\alpha\wedge\mathbf{B}_0.$$

With this substitution, we find that

$$H_{2n} \equiv \int\bar{\mathcal{H}}_{2n}\,d\mathbf{v} = \Gamma_n\hat{E}_n\hat{E}_n^*, \tag{10}$$

$$\int\bar{\mathcal{L}}_3\,d\mathbf{v} = \Gamma_{wc}\hat{E}_1\hat{E}_2\hat{E}_3^* + \mathrm{c.c.}, \tag{11}$$

where the asterisk denotes complex conjugate (c.c.). The expression for the fourth order wave coupling energy per unit volume will be deferred until Sec. 4. The expression of H_2 in terms of the electric field wave amplitude enables the coefficients Γ_n to be calculated directly from the linear dispersion relation for wave n, since with our gauge for the scalar potential, the energy per unit volume of a longitudinal (L), transverse (T) wave is given respectively by

$$H_n^L = \frac{1}{16\pi}\left(\frac{\partial}{\partial\omega}[\omega\epsilon^L(\mathbf{k},\omega)]\right)_{\omega=\omega_n}\left|\hat{E}_n\right|^2,$$

where ω_n is a solution of the dispersion relation $\epsilon^L(\mathbf{k},\omega) = 0$ and

$$H_n^T = \frac{1}{16\pi}\left(\frac{1}{\omega}\frac{\partial}{\partial\omega}[\omega^2\epsilon^T(\mathbf{k},\omega)]\right)_{\omega=\omega_n}\left|\hat{E}_n\right|^2,$$

where ω_n satisfies $\epsilon^T(\mathbf{k},\omega) = c^2k^2/\omega^2$. It follows immediately that

$$\Gamma^L = (16\pi)^{-1}\frac{\partial}{\partial\omega}[\omega\epsilon^L(\mathbf{k},\omega)] \tag{12}$$

and

$$\Gamma^T = (16\pi\omega)^{-1}\frac{\partial}{\partial\omega}[\omega^2\epsilon^T(\mathbf{k},\omega)]. \tag{13}$$

3. THREE-WAVE INTERACTIONS

For three-wave processes, in view of the frequency conservation relation (1), any interaction is possible in which a high frequency wave decays into a low frequency wave together with a wave of the same type as the high frequency one. For example, in a warm isotropic plasma where electron plasma oscillations (L), ion-acoustic waves (IA), and transverse waves (T) exist, any of the following interactions are possible:

(a) $T \rightleftharpoons T + L$, (b) $L \rightleftharpoons L + IA$, (c) $T \rightleftharpoons T + IA$,

(d) $L + L \rightleftharpoons T$, (e) $L + IA \rightleftharpoons T$, (f) $T + IA \rightleftharpoons L$.

It is clear that (d) will only occur for transverse waves whose frequencies are near $2\omega_{pe}$ where ω_{pe} is the elec-

tron plasma frequency, whilst (e) and (f) are possible only if the frequencies of the transverse waves are near ω_{pe}. In a magnetoactive plasma, further types of interaction process can occur, e.g., $T \rightleftarrows T+T$ since a magnetized plasma is capable of supporting a larger variety of wave motions.

The form of the coupling coefficient Γ_{wc} describing the interaction of three arbitrary plasma waves whose frequency and wavevectors satisfy (a) and (b) is obtained from Eq. (5) where, following (9) we write

$$\mathbf{r}_\alpha = \mathbf{r}_{1\alpha} + \mathbf{r}_{2\alpha} + \mathbf{r}_{3\alpha},$$

$$\phi = \phi_1 + \phi_2 + \phi_3,$$

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3.$$

Substituting these expressions into (5), defining $\chi_n = \hat{\mathbf{E}}_n^T / |\hat{\mathbf{E}}_n^T|$, $\hat{\mathbf{S}}_{n\alpha} = D_{\Omega\alpha}\hat{\mathbf{r}}_{n\alpha}$ and using the gauge potential relations $\hat{\phi}_n = i\hat{E}_n^L/k_n$, $\hat{\mathbf{A}}_n = -ic\hat{\mathbf{E}}_n^T/\omega_n$, then the space–time average of $\mathcal{L}_3$ is given by

$$\begin{aligned}\bar{\mathcal{L}}_3 = \sum_\alpha n_{0\alpha} q_\alpha f_{0\alpha} \Big\{ i \Big[&(\hat{\mathbf{r}}_{2\alpha}\cdot\mathbf{k}_1)(\hat{\mathbf{r}}^*_{3\alpha}\cdot\mathbf{k}_1)\Big(\frac{\hat{E}_1^L}{k_1} + \frac{(\mathbf{v}\cdot\chi_1)\hat{E}_1^T}{\omega_1}\Big) \\ &+ (\hat{\mathbf{r}}_{1\alpha}\cdot\mathbf{k}_2)(\hat{\mathbf{r}}^*_{3\alpha}\cdot\mathbf{k}_2)\Big(\frac{\hat{E}_2^L}{k_2} + \frac{(\mathbf{v}\cdot\chi_2)\hat{E}_2^T}{\omega_2}\Big) - (\hat{\mathbf{r}}_{1\alpha}\cdot\mathbf{k}_3) \\ &\times(\hat{\mathbf{r}}_{2\alpha}\cdot\mathbf{k}_3)\Big(\frac{\hat{E}_3^{L*}}{k_3} + \frac{(\mathbf{v}\cdot\chi_3^*)\hat{E}_3^{T*}}{\omega_3}\Big)\Big] + \Big((\hat{\mathbf{r}}_{2\alpha}\cdot\mathbf{k}_1)(\hat{\mathbf{S}}^*_{3\alpha}\cdot\chi_1)\frac{\hat{E}_1^T}{\omega_1} \\ &+ (\hat{\mathbf{r}}^*_{3\alpha}\cdot\mathbf{k}_1)(\hat{\mathbf{S}}_{2\alpha}\cdot\chi_1)\frac{\hat{E}_1^T}{\omega_1} + (\hat{\mathbf{r}}_{1\alpha}\cdot\mathbf{k}_2)(\hat{\mathbf{S}}^*_{3\alpha}\cdot\chi_2)\frac{\hat{E}_2^T}{\omega_2} \\ &+ (\hat{\mathbf{r}}^*_{3\alpha}\cdot\mathbf{k}_2)(\hat{\mathbf{S}}_{1\alpha}\cdot\chi_2)\frac{\hat{E}_2^T}{\omega_2} + (\hat{\mathbf{r}}_{1\alpha}\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_{2\alpha}\cdot\chi_3^*)\frac{\hat{E}_3^{T*}}{\omega_3} \\ &+ (\hat{\mathbf{r}}_{2\alpha}\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_{1\alpha}\cdot\chi_3^*)\frac{\hat{E}_3^{T*}}{\omega_3}\Big) + \text{c.c.} \end{aligned} \tag{14}$$

Use of the Lorentz equation, together with the other small-signal equations of motion, enables $\hat{\mathbf{r}}_{n\alpha}$ to be expressed in terms of $\hat{\mathbf{E}}_n$ so that (14) may be cast into the required form

$$\int \bar{\mathcal{L}}_3 \, d\mathbf{v} = \Gamma_{wc}\hat{E}_1\hat{E}_2\hat{E}_3^* + \text{c.c.} \tag{15}$$

We observe that the effect of the averaging has been to separate synchronous terms, i.e., those which have no net phase dependence, from those with nonzero phase which vanish. The physics of the wave interaction is contained in $\Gamma_{wc} = \Gamma_{wc}(\omega_n, \mathbf{k}_n)$, $n = 1,2,3$, and the space–time evolution of the electric field wave amplitudes of the interacting waves is then prescribed by (c). For any specific interaction, an expression for Γ_{wc} is obtained by inserting the appropriate $\{\hat{\mathbf{r}}_{n\alpha}\}$ into (14), and in Appendix A explicit results for particle displacement vector amplitudes $\hat{\mathbf{r}}_{n\alpha}$ and the corresponding coefficient Γ_n are given for a number of plasma waves.

Although at first sight, $\bar{\mathcal{L}}_3$ contains a large number of terms, many are absent for a particular interaction, e.g., in the case $\omega_3^T = \omega_1^L + \omega_2^T$, then $\hat{E}_2^L = 0 = \hat{E}_3^L$ and $\hat{E}_1^T = 0$. Terms may also vanish if a one- or two-dimensional analysis is performed, rather than the general three-dimensional one presented here. For a plasma in a constant magnetic field $\mathbf{B}_0 = (0,0,B_0)$ referred to a Cartesian triad $0xyz$ where $\mathbf{v} = (v_\perp\cos\psi, v_\perp\sin\psi, v_\parallel)$, the operator $D_{\Omega\alpha} \equiv \partial/\partial t + \mathbf{v}\cdot\nabla + \mathbf{v}\wedge\Omega_\alpha\circ\partial/\partial v = -i(\omega - \mathbf{k}\cdot\mathbf{v}) - \Omega_\alpha\partial/\partial\psi$. Note also that for a warm isotropic plasma $\Omega_\alpha = 0$ while for a cold plasma $f_{0\alpha} = \delta(\mathbf{v})$ and $D_{\Omega\alpha} \equiv -i\omega$.

The form of the coupling coefficient given by (14) and (15) agrees with that obtained by Larsson and Stenflo.[11]

4. FOUR-WAVE INTERACTIONS

To date, theoretical treatments have concentrated mainly on three-wave interactions since in many practical situations, higher order processes produce no measurable effect on plasma behavior. This is no longer always the case, however, with the intense power sources now available and for this reason it is timely to examine the general problem of four-wave interactions in a homogeneous plasma. For example, although two electron plasma oscillations (plasmons) cannot interact to produce a third, conservation of energy and momentum permit four plasmon processes.

For these quadruple processes, the synchronism conditions are

$$\omega_1 + \omega_2 = \omega_3 + \omega_4, \tag{16}$$

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 + \mathbf{k}_4. \tag{17}$$

The coupling coefficients for these four-wave interactions are obtained from the space–time average of $\mathcal{L}_4$ where $\mathcal{L}_4$ is given by (6). From Sec. 2, we recall that these variables are now separated into their individual wave components

$$\mathbf{r}_\alpha = \sum_{n=1}^4 \mathbf{r}_{n\alpha}, \quad \phi = \sum_{n=1}^4 \phi_n, \quad \mathbf{A} = \sum_{n=1}^4 \mathbf{A}_n$$

with each wave variable given by expressions of the form (8). Use of the gauge potential relations then leads after some algebra to the following expression for $\bar{\mathcal{L}}_4$,

$$\begin{aligned}\bar{\mathcal{L}}_4 = \tfrac{1}{32}\sum_\alpha n_{0\alpha} q_\alpha f_{0\alpha}(\mathbf{x},\mathbf{v}) \Big\{ -2\Big[&(\hat{\mathbf{r}}_2\cdot\mathbf{k}_1)(\hat{\mathbf{r}}^*_3\cdot\mathbf{k}_1)(\hat{\mathbf{r}}^*_4\cdot\mathbf{k}_1)\Big(\frac{\hat{E}_1^L}{k_1} + \frac{(\mathbf{v}\cdot\chi_1)\hat{E}_1^T}{\omega_1}\Big) + (\hat{\mathbf{r}}^*_3\cdot\mathbf{k}_2)(\hat{\mathbf{r}}^*_4\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_1\cdot\mathbf{k}_2)\Big(\frac{\hat{E}_2^L}{k_2} + \frac{(\mathbf{v}\cdot\chi_2)\hat{E}_2^T}{\omega_2}\Big) \\ &+ (\hat{\mathbf{r}}^*_4\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_1\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_3)\Big(\frac{\hat{E}_3^{L*}}{k_3} + \frac{(\mathbf{v}\cdot\chi_3^*)\hat{E}_3^{T*}}{\omega_3}\Big) + (\hat{\mathbf{r}}_1\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_4)(\hat{\mathbf{r}}^*_3\cdot\mathbf{k}_4)\Big(\frac{\hat{E}_4^{L*}}{k_4} + \frac{(\mathbf{v}\cdot\chi_4^*)\hat{E}_4^{T*}}{\omega_4}\Big)\Big] + 2i\Big((\hat{\mathbf{r}}_2\cdot\mathbf{k}_1)(\hat{\mathbf{r}}^*_3\cdot\mathbf{k}_1)(\hat{\mathbf{S}}^*_4\cdot\chi_1)\frac{\hat{E}_1^T}{\omega_1} \\ &+ (\hat{\mathbf{r}}_2\cdot\mathbf{k}_1)(\hat{\mathbf{r}}^*_4\cdot\mathbf{k}_1)(\hat{\mathbf{S}}^*_3\cdot\chi_1)\frac{\hat{E}_1^T}{\omega_1} + (\hat{\mathbf{r}}^*_3\cdot\mathbf{k}_1)(\hat{\mathbf{r}}^*_4\cdot\mathbf{k}_1)(\hat{\mathbf{S}}_2\circ\chi_1)\frac{\hat{E}_1^T}{\omega_1} + (\hat{\mathbf{r}}_1\cdot\mathbf{k}_2)(\hat{\mathbf{r}}^*_3\cdot\mathbf{k}_2)(\hat{\mathbf{S}}^*_4\cdot\chi_2)\frac{\hat{E}_2^T}{\omega_2} + (\hat{\mathbf{r}}_1\cdot\mathbf{k}_2)(\hat{\mathbf{r}}^*_4\cdot\mathbf{k}_2)(\hat{\mathbf{S}}^*_3\cdot\chi_2)\frac{\hat{E}_2^T}{\omega_2} \end{aligned}$$

$$+(\hat{\mathbf{r}}_3^*\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_2)(\hat{\mathbf{S}}_1\cdot\boldsymbol{\chi}_2)\frac{\hat{E}_2^T}{\omega_2}-(\hat{\mathbf{r}}_1\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_4^*\cdot\boldsymbol{\chi}_3^*)\frac{\hat{E}_3^{T*}}{\omega_3}-(\hat{\mathbf{r}}_1\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_2\cdot\boldsymbol{\chi}_3^*)\frac{\hat{E}_3^{T*}}{\omega_3}-(\hat{\mathbf{r}}_2\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_1\cdot\boldsymbol{\chi}_3^*)\frac{\hat{E}_3^{T*}}{\omega_3}$$
$$-(\hat{\mathbf{r}}_1\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_4)(\hat{\mathbf{S}}_3^*\cdot\boldsymbol{\chi}_4^*)\frac{\hat{E}_4^{T*}}{\omega_4}-(\hat{\mathbf{r}}_1\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_3^*\cdot\mathbf{k}_4)(\hat{\mathbf{S}}_2\cdot\boldsymbol{\chi}_4^*)\frac{\hat{E}_4^{T*}}{\omega_4}-(\hat{\mathbf{r}}_2\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_3^*\cdot\mathbf{k}_4)(\hat{\mathbf{S}}_1\cdot\boldsymbol{\chi}_4^*)\frac{\hat{E}_4^{T*}}{\omega_4}\Big)-\sum_{i=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)^2(\hat{\mathbf{r}}_i^*\cdot\mathbf{k}_i)\Big(\frac{\hat{E}_i^{L*}}{k_i}$$
$$+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_i^*)\hat{E}_i^{T*}}{\omega_i}\Big)-2\sum_{\substack{i=1\\i\neq j}}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j^*\cdot\mathbf{k}_i)\Big(\frac{\hat{E}_i^{L*}}{k_i}+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_i^*)\hat{E}_i^{T*}}{\omega_i}\Big)-i\sum_{i=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)^2(\hat{\mathbf{S}}_i^*\cdot\boldsymbol{\chi}_i^*)\frac{\hat{E}_i^{T*}}{\omega_i}-2i\sum_{i=1}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_j)$$
$$\times(\hat{\mathbf{r}}_i^*\cdot\mathbf{k}_j)(\hat{\mathbf{S}}_j\cdot\boldsymbol{\chi}_j^*)\frac{\hat{E}_j^{T*}}{\omega_j}-2i\sum_{\substack{i=1\\i\neq j}}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j^*\cdot\mathbf{k}_i)(\hat{\mathbf{S}}_j\cdot\boldsymbol{\chi}_i^*)\frac{\hat{E}_i^{T*}}{\omega_i}-2i\sum_{\substack{i=1\\i\neq j}}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j\cdot\mathbf{k}_i)(\hat{\mathbf{S}}_j^*\cdot\boldsymbol{\chi}_i^*)\frac{\hat{E}_i^{T*}}{\omega_i}\Big\}+\text{c.c.}, \tag{18}$$

where the subscript α has been dropped from the $\hat{\mathbf{r}}_{n\alpha}$ and $\hat{\mathbf{S}}_{n\alpha}$ for brevity. The terms included in the sums over i and over i and j are the "self-energy" terms, of which more will be said later.

Use of the small-signal equations of motion derived from $\mathcal{L}_2$ gives

$$\int\bar{\mathcal{L}}_4\,d\mathbf{v}=\Gamma_{4wc}\hat{E}_1\hat{E}_2\hat{E}_3^*\hat{E}_4^*+\Gamma_{4wc}^*\hat{E}_1^*\hat{E}_2^*\hat{E}_3\hat{E}_4+\sum_{\substack{i=1\\i\leq j}}^{4}\sum_{j=1}^{4}Q_{ij}\hat{E}_i\hat{E}_i^*\hat{E}_j\hat{E}_j^*, \tag{19}$$

where the coefficients Q_{ij} are real functions of ω_n, $\mathbf{k}_n$ $(n=1,2,3,4)$.

An interesting feature of (19) is the appearance of the terms

$$\sum_{\substack{i=1\\i\leq j}}^{4}\sum_{j=1}^{4}Q_{ij}|\hat{E}_i\hat{E}_j|^2.$$

These are the so-called "self-energy" or "forward scattering" terms which arise when particles in the oscillation interact with the potential of the wave itself and are discussed in Fukai, Krishan, and Harris.[12] As will be observed below, these self-energy terms produce no effect on the power flow between the interacting waves. It should be noted that these self-energy terms are analogous to the quadratic Kerr effect terms which arise in the study of four interacting intense electromagnetic waves in nonlinear crystals.[13]

Having obtained the interaction energy per unit volume, we now derive coupled-mode equations describing four-wave interactions for which a formulation in quantum mechanical terms is particularly convenient.

A. Coupled-mode equations

In the quantum mechanical approach to nonlinear plasma wave interactions one considers the plasma consisting of the plasma particles together with quasiparticles which are the individual quanta of the waves. The interaction between waves can then be thought of as effects involving the emission and absorption of certain quasiparticles by others. Introducing the normalized field amplitudes B_n where $\hat{E}_n=(\hbar\omega_n/\Gamma_n)^{1/2}B_n$ where $2\pi\hbar$ is Planck's constant, then from (10) and (19), the nth wave energy density and wave coupling energy density become

$$H_{2n}=\hbar\omega_n B_n^\dagger B_n=\hbar\omega_n N_n, \tag{20}$$

$$\int\bar{\mathcal{L}}_4\,d\mathbf{v}=\hbar^2\Big(\frac{\omega_1\omega_2\omega_3\omega_4}{\Gamma_1\Gamma_2\Gamma_3\Gamma_4}\Big)^{1/2}[\Gamma_{4wc}B_1B_2B_3^\dagger B_4^\dagger+\Gamma_{4wc}^\dagger B_1^\dagger B_2^\dagger \times B_3B_4]+\hbar^2\sum_{\substack{i=1\\i\leq j}}^{4}\sum_{j=1}^{4}\frac{\omega_i\omega_j}{\Gamma_i\Gamma_j}Q_{ij}B_iB_i^\dagger B_jB_j^\dagger, \tag{21}$$

where † denotes Hermitian conjugate. From (20) it can be seen that the energy density for wave n is equivalent to N_n independent quantum oscillators, each with energy $\hbar\omega_n$. The transition from classical to quantum mechanics can then be made by interpreting B_n and $B_n^\dagger$ as annihilation and creation operators respectively for quasiparticles of wave n, so that, apart from sign, (21) now represents the interaction Hamiltonian. This transition from the classical to the quantum mechanical Hamiltonian is the familiar second quantization and it is via this process that the particle nature of the waves arises.

The creation and annihilation operators satisfy the commutation relations for bosons

$$[B_i,B_j]=0,\quad [B_i^\dagger,B_j^\dagger]=0,\quad [B_i,B_j^\dagger]=\delta_{ij},$$

where the commutator $[A,B]=AB-BA$. Physically, we observe that the product of operators appearing in the first term of (21) corresponds to the creation of waves 1 and 2 and the destruction of waves 3 and 4. The second term describes the inverse process. A fuller account of the quantization of waves in dispersive media and the application to nonlinear plasma wave interactions can be found in Askne.[14]

The equations of motion for the B_n follow from the Heisenberg equations

$$i\hbar\frac{dB_n}{dt}=[B_n,H_{\text{int}}]\quad(n=1,2,3,4),$$

where $H_{\text{int}}=-\int\bar{\mathcal{L}}_4\,d\mathbf{v}$ is the interaction Hamiltonian. In order to obtain the equations, the following commutator relations are needed:

$$[B_n,B_1B_2B_3^\dagger B_4^\dagger]=\begin{cases}0 & \text{if } n=1,2,\\ B_1B_2B_4^\dagger & \text{if } n=3,\\ B_1B_2B_3^\dagger & \text{if } n=4,\end{cases}$$

$$[B_n,B_1^\dagger B_2^\dagger B_3B_4]=\begin{cases}B_2^\dagger B_3B_4 & \text{if } n=1,\\ B_1^\dagger B_3B_4 & \text{if } n=2,\\ 0 & \text{if } n=3,4,\end{cases}$$

and

$$[B_n,B_iB_j^\dagger B_kB_l^\dagger]=\delta_{nl}B_iB_j^\dagger B_k+\delta_{nj}B_iB_kB_l^\dagger,$$

where † denotes the Hermitian conjugate and δ_{ij} is the Kronecker delta. The resulting equations obtained after a little algebra are now reinterpreted as classical differential equations for the normalized wave amplitudes. If we reintroduce the electric field wave amplitudes via $B_n = (\Gamma_n/\hbar\omega_n)^{1/2}\hat{E}_n$, define real constants C_{ij} by

$$C_{ii} = 2Q_{ii}, \quad C_{ij} = C_{ji} = Q_{ij} \quad \text{for } i<j, \ i,j = 1,2,3,4,$$

and note that $d/dt \equiv \partial/\partial t + \mathbf{v}_g \cdot \nabla$, these equations produce the required coupled mode equations

$$\begin{aligned}
&\frac{i}{\omega_{1,2}}\left(\frac{\partial}{\partial t} + \mathbf{v}_{g1,2}\cdot\nabla\right)\hat{E}_{1,2} \\
&\quad = -\frac{1}{\Gamma_{1,2}}\left[\Gamma^*_{4wc}\hat{E}^*_{2,1}\hat{E}_3\hat{E}_4 + \hat{E}_{1,2}\sum_{i=1}^{4} C_{1,2i}\hat{E}_i\hat{E}^*_i\right] \\
&\frac{i}{\omega_{3,4}}\left(\frac{\partial}{\partial t} + \mathbf{v}_{g3,4}\cdot\nabla\right)\hat{E}_{3,4} \\
&\quad = -\frac{1}{\Gamma_{3,4}}\left[\Gamma_{4wc}\hat{E}_1\hat{E}_2\hat{E}^*_{4,3} + \hat{E}_{3,4}\sum_{i=1}^{4} C_{3,4i}\hat{E}_i\hat{E}^*_i\right].
\end{aligned} \tag{22}$$

These equations can be cast into canonical form by defining

$$\hat{E}_j = \left(\frac{\omega_j}{\Gamma_j}\right)^{1/2} F_j, \quad \Gamma_{4wc} = iV\left(\frac{\Gamma_1\Gamma_2\Gamma_3\Gamma_4}{\omega_1\omega_2\omega_3\omega_4}\right)^{1/2},$$

$C_{ji} = -i\Gamma_i\Gamma_j g_{ji}/\omega_i\omega_j$ together with the operator

$$D_n = \frac{\partial}{\partial t} + \mathbf{v}_{gn}\cdot\nabla,$$

and (22) becomes

$$\begin{aligned}
D_{1,2}F_{1,2} &= V^*F^*_{2,1}F_3F_4 + F_{1,2}\sum_{i=1}^{4} g_{1,2i}F_iF^*_i \\
D_{3,4}F_{3,4} &= -VF_1F_2F^*_{4,3} + F_{3,4}\sum_{i=1}^{4} g_{3,4i}F_iF^*_i.
\end{aligned} \tag{23}$$

Multiplying these equations by $F^*_{1,2,3,4}$ respectively, adding its complex conjugate and noting that the coefficients g_{ji} are pure imaginary, produces the action transfer relations in the form

$$D_1\left(\frac{\mathcal{E}_1}{\omega_1}\right) = D_2\left(\frac{\mathcal{E}_2}{\omega_2}\right) = -D_3\left(\frac{\mathcal{E}_3}{\omega_3}\right) = -D_4\left(\frac{\mathcal{E}_4}{\omega_4}\right) = \Theta, \tag{24}$$

where $\mathcal{E}_n/\omega_n = (\Gamma_n/\omega_n)|\hat{E}_n|^2$ is the action density of wave n (energy density divided by frequency) and $\Theta = VF_1F_2F^*_3F^*_4 + \text{c.c.}$ From (24) we observe that the self-energy terms which appear in the coupled-mode equations have no effect on the rate of transfer of energy between the waves participating in the interaction.

B. Solutions of the coupled-mode equations for four-wave interactions

We now obtain analytic solutions of the coupled-mode equations (23) and in order to simplify the analysis, we assume that the electric field wave amplitudes vary only with time. Writing

$$F_i(t) = a_i(t)\exp\{-i[\omega_i t + \beta_i(t)]\} \tag{25}$$

with a_i, β_i real and $V = |V|e^{i\delta}$, separation of the real and imaginary parts of (23) gives

$$\begin{aligned}
\frac{\partial a_{1,2}}{\partial t} &= |V| a_{2,1}a_3a_4\cos(\theta+\delta), \\
\frac{\partial a_{3,4}}{\partial t} &= -|V| a_1a_2a_{4,3}\cos(\theta+\delta),
\end{aligned} \tag{26}$$

and

$$\begin{aligned}
-\left(\omega_1 + \frac{\partial\beta_1}{\partial t}\right) &= -|V|\sin(\theta+\delta)\frac{a_2a_3a_4}{a_1} + \sum_{i=1}^{4} h_{1i}a_i^2, \\
-\left(\omega_2 + \frac{\partial\beta_2}{\partial t}\right) &= -|V|\sin(\theta+\delta)\frac{a_1a_3a_4}{a_2} + \sum_{i=1}^{4} h_{2i}a_i^2, \\
\left(\omega_3 + \frac{\partial\beta_3}{\partial t}\right) &= |V|\sin(\theta+\delta)\frac{a_1a_2a_4}{a_3} - \sum_{i=1}^{4} h_{3i}a_i^2, \\
\left(\omega_4 + \frac{\partial\beta_4}{\partial t}\right) &= |V|\sin(\theta+\delta)\frac{a_1a_2a_3}{a_4} - \sum_{i=1}^{4} h_{4i}a_i^2,
\end{aligned}$$

respectively, where $\theta = \beta_4 + \beta_3 - \beta_2 - \beta_1$ and $g_{ji} = ih_{ji}$ so that h_{ji} is real. Adding these last four equations produces an equation for the temporal evolution of the phase difference θ,

$$\begin{aligned}
\frac{\partial\theta}{\partial t} &= -\tan(\theta+\delta)\frac{\partial}{\partial t}\ln[a_1a_2a_3a_4| \\
&\quad + \sum_{i=1}^{4}(h_{1i} + h_{2i} - h_{3i} - h_{4i})a_i^2.
\end{aligned} \tag{27}$$

The Manley–Rowe relations, derived from the action transfer relations (24) are

$$a_4^2 + a_1^2 = \mu_1, \quad a_4^2 + a_2^2 = \mu_2, \quad a_4^2 - a_3^2 = \mu_3 \tag{28}$$

with μ_1, μ_2, and μ_3 constants. Equation (27) can be integrated directly to give

$$\begin{aligned}
&a_1a_2a_3a_4\sin(\theta+\delta) \\
&\quad = \sigma + (4|V|)^{-1}(p_1a_1^4 + p_2a_2^4 - p_3a_3^4 - p_4a_4^4),
\end{aligned} \tag{29}$$

where σ is an integration constant and $p_i = h_{1i} + h_{2i} - h_{3i} - h_{4i}$. From (26)

$$\frac{\partial}{\partial t}a_4^2(t) = -2|V|a_1a_2a_3a_4\cos(\theta+\delta) \tag{30}$$

and using (29), the right-hand side of (30) becomes

$$\begin{aligned}
\frac{\partial}{\partial t}[a_4^2(t)] &= \pm 2|V|\left[(a_1a_2a_3a_4)^2\right. \\
&\quad \left. - \left(\sigma + \frac{1}{4|V|}(p_1a_1^4 + p_2a_2^4 - p_3a_3^4 - p_4a_4^4)\right)^2\right]^{1/2}.
\end{aligned} \tag{31}$$

Using the Manley–Rowe relations to express a_1^2, a_2^2, a_3^2 in terms of a_4^2 leads to the result

$$2|V|t = \pm\int_{a_4^2(0)}^{a_4^2(t)} \frac{d(a_4^2)}{[(a_4^2-\alpha_1)(a_4^2-\alpha_2)(\alpha_3-a_4^2)(\alpha_4-a_4^2)]^{1/2}}, \tag{32}$$

where α_1, α_2, α_3, α_4 are the roots for a_4^2 of $Q=0$ and

$$\begin{aligned}
Q &\equiv a_4^2(a_4^2-\mu_1)(a_4^2-\mu_2)(a_4^2-\mu_3) - \left(\sigma + \frac{1}{4|V|}[p_1(a_4^2-\mu_1)^2\right. \\
&\quad \left. + p_2(a_4^2-\mu_2)^2 - p_3(a_4^2-\mu_3)^2 - p_4a_4^4]\right)^2.
\end{aligned}$$

For bounded solutions corresponding to the transfer of energy continuously among the modes taking part in the interaction, the roots of Q satisfy

$$0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4 \quad \text{with } \alpha_2 \le a_4^2(t) \le \alpha_3.$$

In the general case where $\alpha_{1,2,3,4}$ are all distinct, the solution for $a_4^2(t)$ is given by

$$a_4^2(t) = \frac{\alpha_2(\alpha_3-\alpha_1) - \alpha_1(\alpha_3-\alpha_2)\,\mathrm{sn}^2(\lambda,\gamma)}{(\alpha_3-\alpha_1) - (\alpha_3-\alpha_2)\,\mathrm{sn}^2(\lambda,\gamma)}, \tag{33}$$

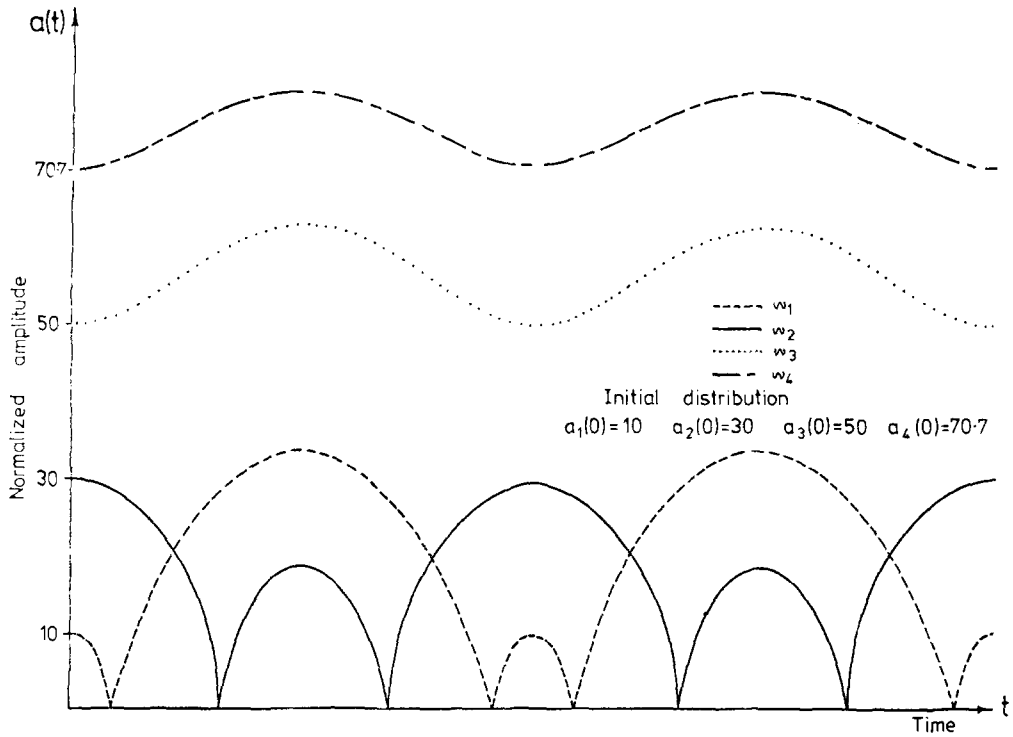


FIG. 1. Plots of the individual wave amplitudes $a_j(t)$ for $\alpha_2/\alpha_1=5$, $\alpha_3/\alpha_1=6$, $\alpha_4/\alpha_1=9$ ($\gamma^2=0.4$), $\alpha_2=5.10^3$, and $a_1(0)=10$, $a_2(0)=30$, $a_3(0)=50$.

where

$$\lambda=|V|[(\alpha_4-\alpha_2)(\alpha_3-\alpha_1)]^{1/2}(t-t_0),$$

$$\gamma=[(\alpha_3-\alpha_2)(\alpha_4-\alpha_1)/(\alpha_4-\alpha_2)(\alpha_3-\alpha_1)]^{1/2}$$

is the modulus of the Jacobian elliptic function sn, and t_0 is a constant defined by

$$\mathrm{sn}[|V|[(\alpha_4-\alpha_2)(\alpha_3-\alpha_1)]^{1/2}t_0,\gamma]=\mp\left[\frac{(\alpha_3-\alpha_1)[a_4^2(0)-\alpha_2]}{(\alpha_3-\alpha_2)[a_4^2(0)-\alpha_1]}\right]^{1/2}.$$

The solutions for the other normalized wave amplitudes are then

$$a_1^2(t)=a_1^2(0)+a_4^2(0)-a_4^2(t),$$

$$a_2^2(t)=a_2^2(0)+a_4^2(0)-a_4^2(t),$$

$$a_3^2(t)=a_3^2(0)-a_4^2(0)+a_4^2(t).$$

The behavior of these wave amplitudes $a_j(t)$, $j=1,2,3,4$, for two sets of initial conditions is shown in Figs. 1 and 2. For convenience we have set $|V|=1$ and $t_0=0$ so that $a_4(0)=\sqrt{\alpha_2}$.

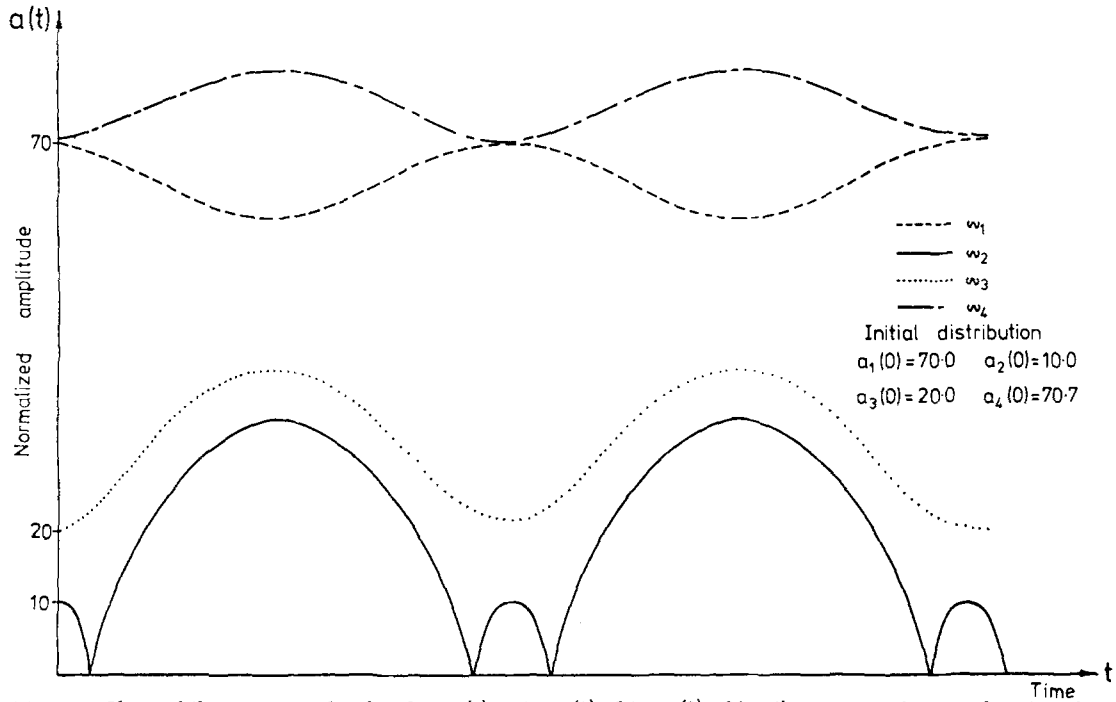


FIG. 2. Plots of the wave amplitudes for $a_1(0)=70$, $a_2(0)=10$, $a_3(0)=20$; other parameters as for Fig. 1.

We conclude this section by noting that the results presented here are readily adapted for the quadruple process where three waves merge into one or one wave splits into three. The synchronism conditions for these processes are

$$\omega_1+\omega_2+\omega_3=\omega_4,$$
$$\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=\mathbf{k}_4,$$

and in Appendix B the corresponding expression for $\bar{L}_4$ is given together with the relevant coupled-mode equations.

5. DISCUSSION

In this paper we have derived general results for the nonlinear interaction of three and four waves in a warm homogeneous plasma. For a particular interaction the $\hat{\mathbf{r}}_{n\alpha}$ for each of the interacting waves is calculated explicitly and inserted into the averaged Lagrangian. The coupling coefficient for the interaction is then obtained directly as described in Sec. 3 and 4 while the coefficients Γ_n which are also required in the coupled-mode equations describing the interaction are derived from the relevant dispersion relation. A detailed, but not exhaustive list of expressions for $\hat{\mathbf{r}}_{n\alpha}$ and Γ_n for a variety of plasma waves in cold and warm plasma is tabulated in Appendix A. As an illustration of the general theory, consider the interaction between four longitudinal waves in a warm isotropic plasma. The synchronism conditions (16) and (17) permit three distinct nonlinear processes: interaction between four electron plasma oscillations (L), between four ion-acoustic waves (IA), and between two L-modes and two IA-modes. From Table I of Appendix A, substituting $\hat{\mathbf{r}}_\alpha=-q_\alpha\hat{E}\mathbf{k}/m_\alpha k\omega'^2$ where $\omega'=\omega-\mathbf{k}\cdot\mathbf{v}$ into (18) gives

$$\bar{L}_4=\tfrac{1}{32}\sum_\alpha n_{0\alpha}\frac{q_\alpha^4}{m_\alpha^3}f_{0\alpha}(\mathbf{v})\left\{\left[\frac{2(\mathbf{k}_2\cdot\mathbf{k}_1)(\mathbf{k}_3\cdot\mathbf{k}_1)(\mathbf{k}_4\cdot\mathbf{k}_1)}{k_1k_2k_3k_4(\omega_2'\omega_3'\omega_4')^2}+\frac{2(\mathbf{k}_3\cdot\mathbf{k}_2)(\mathbf{k}_4\cdot\mathbf{k}_2)(\mathbf{k}_1\cdot\mathbf{k}_2)}{k_1k_2k_3k_4(\omega_3'\omega_4'\omega_1')^2}+\frac{2(\mathbf{k}_4\cdot\mathbf{k}_3)(\mathbf{k}_1\cdot\mathbf{k}_3)(\mathbf{k}_2\cdot\mathbf{k}_3)}{k_1k_2k_3k_4(\omega_4'\omega_1'\omega_2')^2}\right.$$
$$\left.+\frac{2(\mathbf{k}_1\cdot\mathbf{k}_4)(\mathbf{k}_2\cdot\mathbf{k}_4)(\mathbf{k}_3\cdot\mathbf{k}_4)}{k_1k_2k_3k_4(\omega_1'\omega_2'\omega_3')^2}\right]\hat{E}_1\hat{E}_2\hat{E}_3^*\hat{E}_4^*+\sum_{i=1}^{4}\frac{(\mathbf{k}_i\cdot\mathbf{k}_i)^3}{k_i^4\omega_i'^6}(\hat{E}_i\hat{E}_i^*)^2+2\sum_{i=1}^{4}\sum_{\substack{j=1\\ i\neq j}}^{4}\frac{(\mathbf{k}_i\cdot\mathbf{k}_i)(\mathbf{k}_j\cdot\mathbf{k}_i)^2}{k_i^2k_j^2(\omega_i'\omega_j'^2)^2}\hat{E}_i\hat{E}_i^*\hat{E}_j\hat{E}_j^*\right\}+\text{c.c.}$$

For the interaction $\omega_1^L+\omega_2^L=\omega_3^L+\omega_4^L$, (17) implies that $k_1^2+k_2^2=k_3^2+k_4^2$ and the ion contribution to $\bar{L}_4$ may be neglected. Assuming a Maxwellian distribution and noting that $\omega_n'\simeq\omega_n$, the coupling coefficient Γ_{4wc}, and Q_{ii} and Q_{ij} may be written immediately, i.e.,

$$\Gamma_{4wc}=\frac{e^2\omega_{pe}^2}{64\pi m_e^2k_1k_2k_3k_4}\left(\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)(\mathbf{k}_1\cdot\mathbf{k}_3)(\mathbf{k}_1\cdot\mathbf{k}_4)}{(\omega_2\omega_3\omega_4)^2}+\frac{(\mathbf{k}_2\cdot\mathbf{k}_1)(\mathbf{k}_2\cdot\mathbf{k}_3)(\mathbf{k}_2\cdot\mathbf{k}_4)}{(\omega_1\omega_3\omega_4)^2}+\frac{(\mathbf{k}_3\cdot\mathbf{k}_1)(\mathbf{k}_3\cdot\mathbf{k}_2)(\mathbf{k}_3\cdot\mathbf{k}_4)}{(\omega_1\omega_2\omega_4)^2}+\frac{(\mathbf{k}_4\cdot\mathbf{k}_1)(\mathbf{k}_4\cdot\mathbf{k}_2)(\mathbf{k}_4\cdot\mathbf{k}_3)}{(\omega_1\omega_2\omega_3)^2}\right),$$

$$Q_{ii}=\frac{e^2\omega_{pe}^2k_i^2}{64\pi m_e^2\omega_i^6},\qquad Q_{ij}=\frac{e^2\omega_{pe}^2(\mathbf{k}_i\cdot\mathbf{k}_j)^2}{32\pi m_e^2(\omega_i\omega_j)^2}\left(\frac{1}{k_j^2\omega_j^2}+\frac{1}{k_i^2\omega_i^2}\right)\quad(i<j).$$

These results, agree with those obtained by Zakharov[15] for a cold plasma. For the process where four ion-acoustic waves interact, so that $\omega_1^{IA}+\omega_2^{IA}=\omega_3^{IA}+\omega_4^{IA}$, both electron and ion terms are retained. On noting that $\omega_n'\simeq\omega_n$, k_nv_e for the ion, electron contribution respectively, the coupling coefficients become

$$\Gamma_{4wc}=\frac{e^2\omega_{pi}^2}{64\pi m_i^2k_1k_2k_3k_4}\left(\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)(\mathbf{k}_1\cdot\mathbf{k}_3)(\mathbf{k}_1\cdot\mathbf{k}_4)}{(\omega_2\omega_3\omega_4)^2}+\frac{(\mathbf{k}_2\cdot\mathbf{k}_1)(\mathbf{k}_2\cdot\mathbf{k}_3)(\mathbf{k}_2\cdot\mathbf{k}_4)}{(\omega_1\omega_3\omega_4)^2}+\frac{(\mathbf{k}_3\cdot\mathbf{k}_1)(\mathbf{k}_3\cdot\mathbf{k}_2)(\mathbf{k}_3\cdot\mathbf{k}_4)}{(\omega_1\omega_2\omega_4)^2}+\frac{(\mathbf{k}_4\cdot\mathbf{k}_1)(\mathbf{k}_4\cdot\mathbf{k}_2)(\mathbf{k}_4\cdot\mathbf{k}_3)}{(\omega_1\omega_2\omega_3)^2}\right)$$
$$+\frac{e^2\omega_{pe}^2v_e^{-6}}{64\pi m_e^2k_1k_2k_3k_4}\left(\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)(\mathbf{k}_1\cdot\mathbf{k}_3)(\mathbf{k}_1\cdot\mathbf{k}_4)}{(k_2k_3k_4)^2}+\frac{(\mathbf{k}_2\cdot\mathbf{k}_1)(\mathbf{k}_2\cdot\mathbf{k}_3)(\mathbf{k}_2\cdot\mathbf{k}_4)}{(k_1k_3k_4)^2}+\frac{(\mathbf{k}_3\cdot\mathbf{k}_1)(\mathbf{k}_3\cdot\mathbf{k}_2)(\mathbf{k}_3\cdot\mathbf{k}_4)}{(k_1k_2k_4)^2}+\frac{(\mathbf{k}_4\cdot\mathbf{k}_1)(\mathbf{k}_4\cdot\mathbf{k}_2)(\mathbf{k}_4\cdot\mathbf{k}_3)}{(k_1k_2k_3)^2}\right),$$

$$Q_{ii}=\frac{e^2k_i^2}{64\pi}\left\{\frac{\omega_{pi}^2}{m_i^2\omega_i^6}+\frac{\omega_{pe}^2}{m_e^2k_i^6v_e^6}\right\},\quad Q_{ij}=\frac{e^2(\mathbf{k}_i\cdot\mathbf{k}_j)^2}{32\pi}\left\{\frac{\omega_{pi}^2}{m_i^2(\omega_i\omega_j)^2}\left[\frac{1}{k_j^2\omega_j^2}+\frac{1}{k_i^2\omega_i^2}\right]+\frac{\omega_{pe}^2}{m_e^2(k_ik_jv_e^2)^2}\left[\frac{1}{k_j^4v_e^2}+\frac{1}{k_i^4v_e^2}\right]\right\}(i<j).$$

Also from Table I, we note that $\Gamma^L\simeq(8\pi)^{-1}$ and $\Gamma^{IA}\simeq\omega_{pi}^2/8\pi\omega_{IA}^2$. Corresponding expressions for the nonlinear interaction between ion-acoustic and electron plasma oscillations can be similarly obtained.

The theory presented here can be readily extended to include nonlinear coupling in plasmas with particle drift motions, linear damping of the individual modes, frequency and/or wave-vector mismatch and weak inhomogeneity. Wave damping can be taken into account by the transition

$$\frac{\partial}{\partial t}+\mathbf{v}_{gn}\cdot\nabla\to\frac{\partial}{\partial t}+\gamma_n+\mathbf{v}_{gn}\cdot\nabla$$

in Eqs. (3) and (22) where γ_n denotes the linear damping of wave n.

In the case of frequency/wave-vector mismatch, denoting $\mathbf{k}_3-\mathbf{k}_2-\mathbf{k}_1$ by $\Delta\mathbf{k}$ and $\omega_3-\omega_2-\omega_1$ by $\Delta\omega$ for three-wave interactions (with corresponding expressions for quadruple processes), the coupled-mode equations are modified by making the transformations

$$\Gamma_{wc}\to\Gamma_{wc}\exp i(\Delta\omega t-\Delta\mathbf{k}\cdot\mathbf{x}),\quad\Gamma_{4wc}\to\Gamma_{4wc}\exp[i(\Delta\omega t-\Delta\mathbf{k}\cdot\mathbf{x})].$$

Weak inhomogeneity may also be included within the framework of the theory by amending the space–time behavior of the nonlinear wave parameters given by (8). For a spatially inhomogeneous plasma whose properties vary with x, the nonlinear waves are now described by

$$U_n = \mathrm{Re}\{\hat{U}_n(\mathbf{x},t)\exp[i(\mathbf{k}_n\cdot\mathbf{x}-\omega_n t)+i\int_0^x \delta k_n(x)\,dx]\} \tag{34}$$

with $\delta k_n(0)=\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3(-\mathbf{k}_4)=0=\omega_1+\omega_2-\omega_3(-\omega_4)$. Thus at $x=0$, the local frequency and wave vectors satisfy the synchronism conditions in a homogeneous plasma. The quantity $\delta k_n(x)$ is found from the local dispersion relation for wave n in the inhomogeneous plasma. We have shown that the equations describing the evolution of three (four) nonlinearly interacting waves in a weakly inhomogeneous plasma (denoted by I) are given by the coupled-mode equations (3), (22) for the homogeneous plasma (denoted by H), with the inhomogeneous coupling coefficients Γ^I_{wc}, Γ^I_{4wc} being given by

$$\Gamma^I_{wc}\to\Gamma^H_{wc}\exp[i\textstyle\int_0^x K(x)\,dx],\quad \Gamma^I_{4wc}\to\Gamma^H_{4wc}\exp[i\int_0^x K'(x)\,dx],$$

where $K(x)=\delta k_1+\delta k_2-\delta k_3$ and $K'(x)=\delta k_1+\delta k_2-\delta k_3-\delta k_4$. Rosenbluth[16] has examined a three-wave parametric instability in connection with laser plasma interactions for a weakly inhomogeneous plasma using these coupled-mode equations.

APPENDIX A

The particle displacement vector $\mathbf{r}_\alpha$, dispersion relation, and wave polarization are determined from the small-signal equations of motion derived from $\mathcal{L}_2$. Variation of $\mathcal{L}_2$ with respect to $\mathbf{r}_\alpha$ produces the Lorentz equation

$$m_\alpha D^2_{\Omega\alpha}\mathbf{r}_\alpha = q_\alpha\left(\mathbf{E}+\frac{1}{c}\mathbf{v}\wedge\mathbf{B}\right)+\frac{q_\alpha}{c}D_{\Omega\alpha}\mathbf{r}_\alpha\wedge\mathbf{B}_0. \tag{A1}$$

Variation of $\mathcal{L}_2$ with respect to ϕ produces the first-order Poisson equation

$$\nabla\cdot\mathbf{E}=-4\pi\sum_\alpha q_\alpha\int n_{0\alpha}f_{0\alpha}(\nabla\cdot\mathbf{r}_\alpha)\,d\mathbf{v} \tag{A2}$$

whilst variation with respect to $\mathbf{A}$ produces the Maxwell equation

$$\nabla\wedge\mathbf{B}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi}{c}\sum_\alpha q_\alpha\int d\mathbf{v}[n_{0\alpha}f_{0\alpha}(D_{\Omega\alpha}\mathbf{r}_\alpha)-\mathbf{v}\nabla\cdot(n_{0\alpha}f_{0\alpha}\mathbf{r}_\alpha)]. \tag{A3}$$

For longitudinal waves, use of (A1) and (A2) produces the dispersion relation $\epsilon^L(\mathbf{k},\omega)=0$ from which Γ^L can be calculated [see Eq. (12)] while for transverse waves, the dispersion relation and wave polarization for a specific mode are determined by (A1) and (A3).

Solutions of the appropriate equations then produce an expression for the linear wave of the form

$$\mathbf{r}_\alpha=\mathrm{Re}[\hat{\mathbf{r}}_\alpha\exp i(\mathbf{k}\cdot\mathbf{x}-\omega t)]=\mathrm{Re}[F(\omega,\mathbf{k},\mathbf{v})\hat{\mathbf{E}}\exp i(\mathbf{k}\cdot\mathbf{x}-\omega t)] \tag{A4}$$

with $\hat{\mathbf{r}}_\alpha$, $\hat{\mathbf{E}}$ constants in space and time and where the function $F(\omega,\mathbf{k},\mathbf{v})$ is determined by the particular plasma wave. In accordance with the theory presented in Sec. 2 we now assume that the solutions of the nonlinear equations describing wave coupling have the form (A4) with $\hat{\mathbf{r}}_\alpha$, $\hat{\mathbf{E}}$ now being treated as slowly varying functions of $\mathbf{x}$ and t. Hence omitting the exponential factor in (A4), the slowly varying particle displacement vector $\hat{\mathbf{r}}_\alpha$ which will be placed into (14) or (18) can be expressed in the form

$$\hat{\mathbf{r}}_\alpha=F(\omega,\mathbf{k},\mathbf{v})\hat{\mathbf{E}}. \tag{A5}$$

Tables of results for a number of plasma waves are presented and we define the following parameters:

Electron/ion plasma frequency $\omega_{pe,i}=(4\pi n_{0e,i}e^2/m_{e,i})^{1/2}$,

electron/ion temperature $T_{e,i}$,

electron/ion thermal speed $v_{e,i}=(\kappa T_{e,i}/m_{e,i})^{1/2}$,

electron/ion Debye length $\lambda_{De,i}=v_{e,i}/\omega_{pe,i}$,

TABLE I. Warm isotropic plasma.

Mode	Dispersion relation	Γ	$\hat{\mathbf{r}}_\alpha$
Longitudinal	$1-\sum_\alpha\omega^2_{p\alpha}\int\frac{f_{0\alpha}d\mathbf{v}}{(\omega-\mathbf{k}\cdot\mathbf{v})^2}=0$	$\frac{\omega}{8\pi}\sum_\alpha\omega^2_{p\alpha}\int\frac{f_{0\alpha}d\mathbf{v}}{(\omega-\mathbf{k}\cdot\mathbf{v})^3}$	$\frac{-q_\alpha\hat{\mathbf{E}}}{m_\alpha(\omega-\mathbf{k}\cdot\mathbf{v})^2}$
Electron plasma oscillation	$\omega^2\simeq\omega^2_{pe}+3k^2v_e^2$	$(8\pi)^{-1}$	$\hat{\mathbf{r}}_i=0$ $\hat{\mathbf{r}}_e=e\hat{\mathbf{E}}/m_e\omega^2$
Ion-acoustic wave	$\omega^2\simeq\frac{k^2c_s^2}{1+k^2\lambda^2_{De}}$	$\frac{\omega^2_{pi}}{8\pi\omega^2}$	$\hat{\mathbf{r}}_i=-e\hat{\mathbf{E}}/m_i\omega^2$ $\hat{\mathbf{r}}_e=e\hat{\mathbf{E}}/m_ek^2v_e^2$
Transverse electromagnetic wave	$1-\frac{\omega^2_{pe}}{\omega^2}=\frac{c^2k^2}{\omega^2}$	$(8\pi)^{-1}$	$\hat{\mathbf{r}}_i=0$; $\hat{\mathbf{r}}_e=e\hat{\mathbf{E}}/m_e\omega^2$

TABLE II. Cold anisotropic plasma—longitudinal waves.

Mode	Dispersion relation	Γ	$\hat{\mathbf{r}}_\alpha$
Upper-hybrid wave	$\omega^2 \simeq \frac{[\omega_{UH}^2 + (\omega_{UH}^4 - 4\omega_{pe}^2\Omega_e^2\cos^2\theta)^{1/2}]}{2}$	$\frac{1}{8\pi}\left(1+\frac{\omega^2{}_{pe}\Omega_e^2\sin^2\theta}{(\omega^2-\Omega_e^2)^2}\right)$	$\hat{\mathbf{r}}_i = 0$ $\hat{\mathbf{r}}_e = \frac{e\hat{E}}{m_e k(\omega^2-\Omega_e^2)}\left(k_x + \frac{i\Omega_e}{\omega}k_y, k_y - \frac{i\Omega_e}{\omega}k_x, \frac{\omega^2-\Omega_e^2}{\omega^2}k_z\right)$
Electron plasma wave	$\omega^2 \simeq \frac{\omega_{pe}^2\Omega_e^2\cos^2\theta}{\omega_{UH}^2}$	as above	as above
Lower-hybrid wave	$\omega^2 \simeq \omega_{LH}^2\left(1+\frac{m_i}{m_e}\cos^2\theta\right)$	$\frac{1}{8\pi}\left(1+\frac{\omega_{pe}^2\Omega_e^2\sin^2\theta}{(\omega^2-\Omega_e^2)^2}\right)$	$\hat{\mathbf{r}}_\alpha = \frac{-q_\alpha\hat{E}}{m_\alpha k(\omega^2-\Omega_\alpha^2)}\left(k_x + \frac{i\Omega_\alpha}{\omega}k_y, k_y - \frac{i\Omega_\alpha}{\omega}k_x, \frac{\omega^2-\Omega_\alpha^2}{\omega^2}k_z\right)$

upper-hybrid frequency $\omega_{UH} = (\omega_{pe}^2 + \Omega_e^2)^{1/2}$,

lower-hybrid frequency $\omega_{LH} = \omega_{pe}(|\Omega_e|\Omega_i)^{1/2}/\omega_{UH}$,

Alfvén speed $v_A = B_0/(4\pi n_{0e}m_i)^{1/2}$,

ion-acoustic speed $c_s = (m_e/m_i)^{1/2}v_e$.

We take $\mathbf{B}_0 = (0, 0, B_0)$ referred to a Cartesian frame $0xyz$ with B_0 constant and $\theta = \cos^{-1}(\mathbf{k}\cdot\hat{\mathbf{z}}/k)$. We also define $\mathbf{k} = (k_\perp\cos\beta, k_\perp\sin\beta, k_\parallel)$, $\mathbf{v} = (v_\perp\cos\psi, v_\perp\sin\psi, v_\parallel)$;

$$\nu_{L,R}^2 = 2\omega^2(\omega^2-\omega_{pe}^2) - \omega^2\Omega_e^2\sin^2\theta \pm 2\omega|\Omega_e|(\omega^2-\omega_{pe}^2)\cos\theta,$$

$$d^2 = \frac{\omega_{pe}^4}{\omega^4}\left(1+\frac{(\Omega_e^2/\omega^2)[1-(\omega_{pe}^2/\omega^2)\cos^2\theta]^2}{\{(\Omega_e^2\sin^2\theta/\omega^2)+[1-(\omega_{pe}^2/\omega^2)\cos^2\theta]\}^2}\right)$$

$$b_1 = \frac{\omega_{pe}^2(\omega^2-\omega_{pe}^2)\cos^2\theta\cot\theta}{\omega^2[\Omega_e^2\sin^2\theta+(\omega^2-\omega_{pe}^2)\cos^2\theta]}, \quad b_2 = \frac{|\Omega_e|\omega_{pe}^2\sin^2\theta\cot\theta}{\omega[\Omega_e^2\sin^2\theta+(\omega^2-\omega_{pe}^2)\cos^2\theta]}, \quad g = \frac{[\omega(\omega^2-\omega_{pe}^2)-\omega\Omega_e^2\sin^2\theta]}{|\Omega_e|(\omega_{pe}^2-\omega^2\cos^2\theta)},$$

TABLE III. Cold anisotropic plasma—general waves.

Mode	Dispersion relation	Γ	$\hat{\mathbf{r}}_\alpha$
Fast Alfvén wave	$\frac{c^2k^2}{\omega^2} = 1+\frac{c^2}{v_A^2}$	$\frac{1}{8\pi}\left(1+\frac{c^2}{v_A^2}\right)$	$\hat{\mathbf{r}}_\alpha = \frac{-q_\alpha\hat{E}(1-\Omega_\alpha/\omega)}{\sqrt{2}m_\alpha(\omega^2-\Omega_\alpha^2)}(1, -i, 0)$
Slow Alfvén wave	$\frac{c^2k^2\cos^2\theta}{\omega^2} = 1+\frac{c^2}{v_A^2}$	$\frac{1}{8\pi}\left(1+\frac{c^2}{v_A^2}\right)\sec^2\theta$	$\hat{\mathbf{r}}_\alpha = \frac{-q_\alpha\hat{E}(1+\Omega_\alpha/\omega)}{\sqrt{2}m_\alpha(\omega^2-\Omega_\alpha^2}(1, -i, 0)$
Ion cyclotron wave	$\frac{c^2k^2}{\omega^2} \simeq \frac{2\omega_{pi}^2}{\Omega_i^2-\omega^2}$	$\frac{c^4k^4}{16\pi\omega_{pi}^2\Omega_i^2}$	as above
Whistler or helicon wave	$\frac{c^2k^2}{\omega^2} \simeq \frac{\omega_{pe}^2}{\omega(\lvert\Omega_e\rvert\cos\theta-\omega)}$	$\frac{\omega_{pe}^2\lvert\Omega_e\rvert\cos\theta}{16\pi\omega(\lvert\Omega_e\rvert\cos\theta-\omega)^2}$	$\hat{\mathbf{r}}_i = 0$; $\hat{\mathbf{r}}_e = \frac{e\hat{E}}{m_e\omega(\omega\cos\theta-\lvert\Omega_e\rvert)(1+b^2)^{1/2}}(\cos\theta, i, 0)$, where $b = (\omega-\lvert\Omega_e\rvert\cos\theta)/(\omega\cos\theta-\lvert\Omega_e\rvert)$
Electron cyclotron wave	$\frac{c^2k^2}{\omega^2} \simeq 1-\frac{\omega_{pe}^2}{\omega(\omega-\lvert\Omega_e\rvert\cos\theta)}$	$\frac{1}{16\pi}\left(2+\frac{\omega_{pe}^2\lvert\Omega_e\rvert\cos\theta}{\omega(\omega-\lvert\Omega_e\rvert\cos\theta)^2}\right)$	as above
LCP/RCP	$\epsilon_{L,R}^T = c^2k^2/\omega^2$ where		$\hat{\mathbf{r}}_i = 0$
Electron mode	$\epsilon_{L,R}^T = 1-\frac{2\omega_{pe}^2(\omega^2-\omega_{pe}^2)}{\nu_{L,R}^2}$	$\frac{1}{16\pi\omega}\frac{\partial}{\partial\omega}(\omega^2\epsilon_{L,R}^T)$	$\hat{\mathbf{r}}_{eL,R} = \frac{e\hat{E}}{\sqrt{2}m_e(\omega^2-\Omega_e^2)}\left(1\mp\frac{i\lvert\Omega_e\rvert}{\omega}, \mp i-\frac{\lvert\Omega_e\rvert}{\omega}, 0\right)$
Ordinary mode	$\frac{c^2k^2}{\omega^2} = \frac{\omega^2-\omega_{pe}^2}{\omega^2-\omega_{pe}^2\cos^2\theta}$	$\frac{1}{8\pi}\left(1-\frac{\omega_{pe}^4\sin^2\theta\cos^2\theta}{(\omega^2-\omega_{pe}^2\cos^2\theta)^2}\right)$	$\hat{\mathbf{r}}_i = 0$ $\hat{\mathbf{r}}_e = \frac{e\hat{E}}{m_e\omega^2(1+d^2)^{1/2}}(b_1, -ib_2, 1)$
Extraordinary mode	$\frac{c^2k^2}{\omega^2} = \frac{(1-\omega_{pe}^2/\omega^2)^2-(\Omega_e^2/\omega^2)\sin^2\theta}{(1-\omega_{pe}^2/\omega^2)-(\Omega_e^2/\omega^2)\sin^2\theta}$		$\hat{\mathbf{r}}_i = 0$
		$\frac{1}{16\pi\omega}\frac{\partial}{\partial\omega}\left(\frac{(\omega^2-\omega_{pe}^2)^2-\omega^2\Omega_e^2\sin^2\theta}{(\omega^2-\omega_{pe}^2)-\Omega_e^2\sin^2\theta}\right)$	$\hat{\mathbf{r}}_e = \frac{e\hat{E}(1+g^2)^{-1/2}}{m_e(\omega_{pe}^2-\omega^2\cos^2\theta)}(\sin^2\theta, ih, 0)$

TABLE IV. Warm anisotropic plasma—Longitudinal waves.

Mode	Dispersion relation	Γ	$\hat{\mathbf{r}}_\alpha$
	$\epsilon^L(\mathbf{k},\omega)=0$ where		$\hat{\mathbf{r}}_\alpha=\dfrac{-q_\alpha\hat{E}e^{i\mu_\alpha\sin(\psi-\beta)}}{m_\alpha k}\sum_{n=-\infty}^{\infty}\dfrac{J_n(\mu_\alpha)e^{-in(\psi-\beta)}}{Q_n}$
Bernstein modes	$\epsilon^L=1-\left\{\sum_\alpha(k\lambda_{D\alpha})^{-2}\int_{-\infty}^{\infty}dv_{\|}f_{0\alpha}(v_{\|})\times\sum_{n=-\infty}^{\infty}\dfrac{(k_{\|}v_{\|}+n\Omega_\alpha)\sigma_{n\alpha}}{(\omega-k_{\|}v_{\|}-n\Omega_\alpha)}\right\}$	$\dfrac{1}{16\pi}\dfrac{\partial}{\partial\omega}(\omega\epsilon^L)$	$\times\left[\dfrac{k_\perp}{2}\left[\dfrac{e^{i\beta}}{Q_{n+1}}+\dfrac{e^{-i\beta}}{Q_{n-1}}\right],\dfrac{k_\perp}{2i}\left[\dfrac{e^{i\beta}}{Q_{n+1}}-\dfrac{e^{-i\beta}}{Q_{n-1}}\right],\dfrac{k_{\|}}{Q_n}\right]$
Cyclotron harmonic waves ($k_{\|}=0$)	$\omega-n\lvert\Omega_\alpha\rvert\simeq\left[\dfrac{\omega_{p\alpha}}{\Omega_\alpha}\right]^2\dfrac{\sigma_{n\alpha}}{\lambda_\alpha}n\lvert\Omega_\alpha\rvert$ $(n\geq 2)$	$\left[\dfrac{\Omega_\alpha}{\omega_{p\alpha}}\right]^2\dfrac{\lambda_\alpha}{16\pi\sigma_{n\alpha}}$	
Electron and ion cyclotron waves ($k_{\|}=0$)	$\omega^2\simeq\Omega_\alpha^2\left\{1+2\left[\dfrac{\omega_{p\alpha}}{\Omega_\alpha}\right]^2\dfrac{\sigma_{1\alpha}}{\lambda_\alpha}\right\}$	$\left[\dfrac{\Omega_\alpha}{\omega_{p\alpha}}\right]^2\dfrac{\lambda_\alpha}{16\pi\sigma_{1\alpha}}$	

$h=(\omega^2-\omega_{pe}^2)/\omega\lvert\Omega_e\rvert$, $\mu_\alpha=k_\perp v_\perp/\Omega_\alpha$, $\lambda_\alpha=k_\perp^2v_\alpha^2/\Omega_\alpha^2$. $Q_p=(\omega-k_{\|}v_{\|}-p\Omega_\alpha)$, $\sigma_{n\alpha}=\exp(-\lambda_\alpha)I_n(\lambda_\alpha)$,

and J_n, I_n are Bessel functions of the first kind.

APPENDIX B

The nonlinear process where one wave decays into three, and three waves coalesce to produce one are described by the synchronism conditions

$$\omega_1+\omega_2+\omega_3=\omega_4,\quad \mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=\mathbf{k}_4.$$

The space-time averaged Lagrangian density describing the interaction is given by

$$\begin{aligned}\bar{\mathcal{L}}_4=\tfrac{1}{32}\sum_\alpha n_{0\alpha}f_{0\alpha}(\mathbf{x},\mathbf{v})q_\alpha\Bigg\{&-2\Bigg[(\hat{\mathbf{r}}_2\cdot\mathbf{k}_1)(\hat{\mathbf{r}}_3\cdot\mathbf{k}_1)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_1)\left(\frac{\hat{E}_1^L}{k_1}+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_1)\hat{E}_1^T}{\omega_1}\right)+(\hat{\mathbf{r}}_3\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_1\cdot\mathbf{k}_2)\left(\frac{\hat{E}_2^L}{k_2}+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_2)\hat{E}_2^T}{\omega_2}\right)\\&+(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_1\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_3)\left(\frac{\hat{E}_3^L}{k_3}+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_3)\hat{E}_3^T}{\omega_3}\right)+(\hat{\mathbf{r}}_1\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_3\cdot\mathbf{k}_4)\left(\frac{\hat{E}_4^{L*}}{k_4}+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_4^*)\hat{E}_4^{T*}}{\omega_4}\right)\Bigg]+2i\Bigg[(\hat{\mathbf{r}}_2\cdot\mathbf{k}_1)(\hat{\mathbf{r}}_3\cdot\mathbf{k}_1)(\hat{\mathbf{S}}_4^*\cdot\boldsymbol{\chi}_1)\frac{\hat{E}_1^T}{\omega_1}\\&+(\hat{\mathbf{r}}_2\cdot\mathbf{k}_1)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_1)(\hat{\mathbf{S}}_3\cdot\boldsymbol{\chi}_1)\frac{\hat{E}_1^T}{\omega_1}+(\hat{\mathbf{r}}_3\cdot\mathbf{k}_1)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_1)(\hat{\mathbf{S}}_2\cdot\boldsymbol{\chi}_1)\frac{\hat{E}_1^T}{\omega_1}+(\hat{\mathbf{r}}_1\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_3\cdot\mathbf{k}_2)(\hat{\mathbf{S}}_4^*\cdot\boldsymbol{\chi}_2)\frac{\hat{E}_2^T}{\omega_2}+(\hat{\mathbf{r}}_1\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_2)(\hat{\mathbf{S}}_3\cdot\boldsymbol{\chi}_2)\frac{\hat{E}_2^T}{\omega_2}\\&+(\hat{\mathbf{r}}_3\cdot\mathbf{k}_2)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_2)(\hat{\mathbf{S}}_1\cdot\boldsymbol{\chi}_2)\frac{\hat{E}_2^T}{\omega_2}+(\hat{\mathbf{r}}_1\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_4^*\cdot\boldsymbol{\chi}_3)\frac{\hat{E}_3^T}{\omega_3}+(\hat{\mathbf{r}}_1\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_2\cdot\boldsymbol{\chi}_3)\frac{\hat{E}_3^T}{\omega_3}+(\hat{\mathbf{r}}_2\cdot\mathbf{k}_3)(\hat{\mathbf{r}}_4^*\cdot\mathbf{k}_3)(\hat{\mathbf{S}}_1\cdot\boldsymbol{\chi}_3)\frac{\hat{E}_3^T}{\omega_3}\\&-(\hat{\mathbf{r}}_1\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_2\cdot\mathbf{k}_4)(\hat{\mathbf{S}}_3\cdot\boldsymbol{\chi}_4^*)\frac{\hat{E}_4^{T*}}{\omega_4}-(\hat{\mathbf{r}}_1\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_3\cdot\mathbf{k}_4)(\hat{\mathbf{S}}_2\cdot\boldsymbol{\chi}_4^*)\frac{\hat{E}_4^{T*}}{\omega_4}-(\hat{\mathbf{r}}_2\cdot\mathbf{k}_4)(\hat{\mathbf{r}}_3\cdot\mathbf{k}_4)(\hat{\mathbf{S}}_1\cdot\boldsymbol{\chi}_4^*)\frac{\hat{E}_4^{T*}}{\omega_4}\Bigg]-\sum_{i=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)^2(\hat{\mathbf{r}}_i^*\cdot\mathbf{k}_i)\Bigg[\frac{\hat{E}_i^{L*}}{k_i}\\&+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_i^*)\hat{E}_i^{T*}}{\omega_i}\Bigg]-2\sum_{\substack{i=1\\i\neq j}}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j^*\cdot\mathbf{k}_i)\Bigg[\frac{\hat{E}_i^{L*}}{k_i}+\frac{(\mathbf{v}\cdot\boldsymbol{\chi}_i^*)\hat{E}_i^{T*}}{\omega_i}\Bigg]-i\sum_{i=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)^2(\hat{\mathbf{S}}_i^*\cdot\boldsymbol{\chi}_i^*)\frac{\hat{E}_i^{T*}}{\omega_i}-2i\sum_{i=1}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_j)\\&\times(\hat{\mathbf{r}}_i^*\cdot\mathbf{k}_j)(\hat{\mathbf{S}}_j\cdot\boldsymbol{\chi}_j^*)\frac{\hat{E}_j^{T*}}{\omega_j}-2i\sum_{\substack{i=1\\i\neq j}}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j^*\cdot\mathbf{k}_i)(\hat{\mathbf{S}}_j\cdot\boldsymbol{\chi}_i^*)\frac{\hat{E}_i^{T*}}{\omega_i}-2i\sum_{\substack{i=1\\i\neq j}}^{4}\sum_{j=1}^{4}(\hat{\mathbf{r}}_i\cdot\mathbf{k}_i)(\hat{\mathbf{r}}_j\cdot\mathbf{k}_i)(\hat{\mathbf{S}}_j^*\cdot\boldsymbol{\chi}_i^*)\frac{\hat{E}_i^{T*}}{\omega_i}\Bigg\}+\text{c.c.}\end{aligned}$$

The self-energy terms are identical for the two types of quadruple process. Using the small-signal equations of motion, the integrated Lagrangian density will be of the form [cf. (19)]

$$\int\bar{\mathcal{L}}_4\,d\mathbf{v}=\gamma_{4wc}\hat{E}_1\hat{E}_2\hat{E}_3\hat{E}_4^*+\gamma_{4wc}^*\hat{E}_1^*\hat{E}_2^*\hat{E}_3^*\hat{E}_4+\sum_{\substack{i=1\\i\leq j}}^{4}\sum_{j=1}^{4}Q_{ij}\hat{E}_i\hat{E}_i^*\hat{E}_j\hat{E}_j^*$$

and the corresponding coupled-mode equations are

$$i\omega_1^{-1}\left(\frac{\partial}{\partial t}+\mathbf{v}_{g1}\cdot\nabla\right)\hat{E}_1=-\Gamma_1^{-1}\left[\gamma_{4wc}^*\hat{E}_2^*\hat{E}_3^*\hat{E}_4+\hat{E}_1\sum_{i=1}^{4}C_{1i}\hat{E}_i\hat{E}_i^*\right],\quad i\omega_2^{-1}\left(\frac{\partial}{\partial t}+\mathbf{v}_{g2}\cdot\nabla\right)\hat{E}_2=-\Gamma_2^{-1}\left[\gamma_{4wc}^*\hat{E}_3^*\hat{E}_1^*\hat{E}_4+\hat{E}_2\sum_{i=1}^{4}C_{2i}\hat{E}_i\hat{E}_i^*\right],$$

$$i\omega_3^{-1}\left(\frac{\partial}{\partial t}+\mathbf{v}_{g3}\cdot\nabla\right)\hat{E}_3=-\Gamma_3^{-1}\left[\gamma_{4wc}^*\hat{E}_1^*\hat{E}_2^*\hat{E}_4+\hat{E}_3\sum_{i=1}^{4}C_{3i}\hat{E}_i\hat{E}_i^*\right],\quad i\omega_4^{-1}\left(\frac{\partial}{\partial t}+\mathbf{v}_{g4}\cdot\nabla\right)\hat{E}_4=-\Gamma_4^{-1}\left[\gamma_{4wc}\hat{E}_1\hat{E}_2\hat{E}_3+\hat{E}_4\sum_{i=1}^{4}C_{4i}\hat{E}_i\hat{E}_i^*\right].$$

[1]P.A. Sturrock, Proc. Roy. Soc. A 242, 277 (1957).
[2]D.F. Dubois, *Laser Interaction and Related Plasma Phenomena*, edited by H.J. Schwarz and H. Hora (Plenum, New York, 1974), Vol. 3A, p. 267.
[3]F.E. Low, Proc. Roy. Soc. A 248, 282 (1958).
[4]R.C. Davidson, *Methods in Nonlinear Plasma Theory* (Academic, New York, 1972).
[5]G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[6]T.J.M. Boyd and J.G. Turner, J. Phys. A Gen. Phys. 5, 881 (1972).
[7]T.J.M. Boyd and J.G. Turner, J. Phys. A Gen. Phys. 6, 272 (1973).
[8]J.P. Dougherty, J. Plasma Phys. 4, 761 (1970).
[9]J.P. Dougherty, J. Plasma Phys. 11, 331 (1974).
[10]N. Krylov and N. Bogoliubov, *Introduction to Nonlinear Mechanics* (Princeton U.P., Princeton, N.J., 1949).
[11]J. Larsson and L. Stenflo, Beit. aus der Plasmaphysik 16, 79 (1976).
[12]J. Fukai, S. Krishan, and E.G. Harris, Phys. Fluids 13, 3031 (1970).
[13]J.A. Armstrong, N. Bloembergen, J. Ducuing, and P.S. Pershan, Phys. Rev. 127, 1918 (1962).
[14]J. Askne, J. Phys. A Gen. Phys. 5, 1578 (1972).
[15]V.E. Zakharov, Sov. Phys. JETP 24, 445 (1967).
[16]M.N. Rosenbluth, Phys. Rev. Lett. 29, 565 (1972).

Path integrals for solving some electromagnetic edge diffraction problems[a)]

S. W. Lee

Electromagnetics Laboratory, Department of Electrical Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801
(Received 8 November 1977)

Electromagnetic edge diffraction problems involving parallel half-planes are traditionally attacked by the Wiener–Hopf technique, or asymptotically for a large wavenumber ($k \to \infty$) by ray-optic techniques. This paper reports a novel method in which the electromagnetic wave equation is first converted to a heat equation via the Laplace transform. The heat equation together with the original boundary condition is next solved approximately in terms of a path integral over the Wiener measure. For several examples involving two parallel half-planes, the path integral is evaluated explicitly to yield an asymptotic solution of order k^0 for the field on the incident shadow boundary. Those solutions agree with the ones derived by traditional techniques, but are obtained here in a much simpler manner. In other examples involving multiple half-planes, the use of a path integral leads to new solutions. We have not succeeded, however, in generating higher-order terms beyond k^0 in the asymptotic solution by path integrals.

1. INTRODUCTION

Path integrals were introduced independently by N. Wiener in 1924 and by R. P. Feynman in 1948. They have since been used extensively in quantum mechanics and studied rigorously as a new branch of mathematics. An introduction to path integrals can be found in Ref. 1 and in the first three chapters of Ref. 2.

In the present paper, the path integral technique is applied to the calculation of the time-harmonic electromagnetic field on the incident shadow boundary in several edge-diffraction problems involving parallel half-planes. The solutions thus obtained are asymptotically valid for a large wavenumber k (2π/wavelength) and include only the dominant terms of order k^0 relative to the incident field. Those solutions do agree, when available, with the exact asymptotic ones derived by analytical techniques[3–7] and/or by a uniform asymptotic theory of ray-optic techniques.[8] The latter techniques involve complicated mathematical manipulations. In contrast, the path integrals yield the asymptotic solutions in a few elementary steps.

All of the edge-diffraction problems treated in this paper depend only on two spatial variables (x, z). Consequently, the complete electromagnetic fields are derivable from a scalar $u(x, z)$, which satisfies a scalar wave equation. Following Buslaev,[9] we convert the problem of solving u to that of solving $K(x, z, t)$, which satisfies a heat equation. The asymptotic solution of u as $k \to \infty$ is related to that of K as $t \to 0$, as shown in Sec. 2. Next, the heat equation for K together with the appropriate boundary and initial conditions is approximately solved in Sec. 3, with its solution expressed in terms of a path integral. In the next three sections, this path integral is evaluated for examples, and their results compared with rigorous ones whenever the latter are available. Finally, a conclusion is given in Sec. 7.

2. HEAT EQUATION FORMULATION

To introduce the path integral, consider the two-dimensional diffraction problem in Fig. 1. A conducting half-plane Σ located at $x < a \sin\theta$, $z = a\cos\theta$ is illuminated by the incident field from a time-harmonic line source at Q where $(x = 0,\ z = 0)$. The problem is to determine the total field at an observation point B which is located exactly on the incident shadow boundary AB, and has coordinates $[x = (a + b)\sin\theta,\ z = (a + b)\times\cos\theta]$. The incident as well as the total fields may be resolved into two modes: E wave (with nonzero field components E_y, H_x, and H_z) and H wave (H_y, E_x, and E_z). We will treat both modes simultaneously with the help of two notations:

$$\text{for } E \text{ wave: } u = E_y,\ R = -1, \tag{2.1a}$$

$$\text{for } H \text{ wave: } u = H_y,\ R = +1. \tag{2.1b}$$

It is convenient to associate R with the reflection coefficient of u from a conducting plane. All of the fields have the same temporal variation $\exp(-i\omega T)$ as the source, and this common factor will be omitted throughout this paper. The present boundary value problem can be exactly formulated in terms of the wave equation for the total field u

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\right) u(x, z) = -\delta(x)\delta(z), \tag{2.2}$$

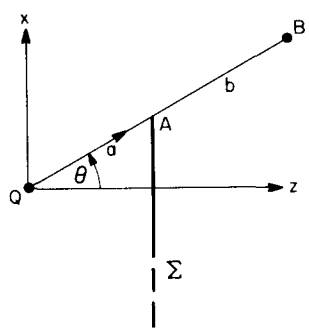


FIG. 1. A half-plane Σ illuminated by the incident field from a line source at Q.

[a)] This work is supported by National Science Foundation Grant ENG73-08218.

 0022-2488/78/1906-1414$1.00

the boundary condition

$$\left.\begin{array}{l}\text{for } E \text{ wave: } u=0\\ \text{for } H \text{ wave: } \partial u/\partial z=0\end{array}\right\}, \quad \text{for } (x,z) \text{ on } \Sigma, \tag{2.3}$$

and the radiation condition. In (2.2), $k=\omega/c$ is the wavenumber and $\delta(\cdot)$ is a Dirac delta function. The radiation condition can be enforced by allowing k to be slightly complex with

$$\mathrm{Re}k>0, \quad \mathrm{Im}k>0 \tag{2.4}$$

and requiring that u decays exponentially as $(x^2+z^2)^{1/2} \to \infty$. As formulated above, this problem can be solved exactly by the Wiener–Hopf or other analytical techniques, with its solution given in terms of two integrals [see Sec. 11.7, Eq. (20) of Ref. 10]. At high frequencies, these integrals can be asymptotically evaluated to yield the total field at B:

$$u(B)=\tfrac{1}{2}u^i(B)+R\sec\theta\, g(kb)u^i(A)+O(k^{-2}), \quad k\to\infty, \tag{2.5}$$

valid for θ away from $\pi/2$. In (2.5), $u(B)$, for example, means $u(x,z)$ evaluated at point B, and g is a cylindrical wave factor

$$g(kx)=(8\pi kx)^{-1/2}\exp[i(kx+\pi/4)]. \tag{2.6}$$

The incident field u^i in (2.5), which is identified with the field radiated from the line source in the absence of Σ, is given by

$$u^i(x,z)=(i/4)H_0^{(1)}[k(x^2+z^2)^{1/2}] \tag{2.7a}$$

$$=g[k(x^2+z^2)^{1/2}]+O(k^{-3/2}), \quad k\to\infty. \tag{2.7b}$$

The solution in (2.5) can be also obtained by the uniform asymptotic theory,[11] a refinement of Keller's geometrical theory of diffraction.[12]

Now for the same diffraction problem sketched in Fig. 1, let us consider a different formulation which will be later solved by a path integral in Sec. 3. This formulation, described first by Buslaev,[9] is related to the procedure for solving the time-independent Schrödinger equation (see Sec. 6 of Ref. 1). Consider the auxiliary problem for an unknown field $K(x,z,t)$, which is zero for $t<0$ and satisfies the heat (diffusion) equation

$$\frac{\partial}{\partial t}K=\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)K, \quad \text{for } t>0. \tag{2.8}$$

Furthermore, K satisfies the same boundary condition as u:

$$\left.\begin{array}{l}\text{for } E \text{ wave: } K=0\\ \text{for } H \text{ wave: } \partial K/\partial z=0\end{array}\right\}, \quad \text{for } (x,z) \text{ on } \Sigma, \tag{2.9}$$

and an initial condition

$$K(x,z,t=0+)=\delta(x)\delta(z). \tag{2.10}$$

To relate K to u, let us take the Laplace transform of (2.8) and obtain

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}-s\right)\hat{K}(x,z,s)=-\delta(x)\delta(z), \tag{2.11}$$

where

$$\hat{K}(x,z,s)=\int_0^\infty K(x,z,t)e^{-st}dt, \quad \mathrm{Re}s>0. \tag{2.12}$$

A comparison of (2.11) with (2.2) leads to the conclusion that $u=\hat{K}$ after replacing s by $-k^2$, or

$$u(x,z)=\int_0^\infty K(x,z,t)e^{k^2t}dt, \quad \pi/4<\arg k<3\pi/4. \tag{2.13}$$

Thus, the diffraction problem in Fig. 1 has been converted to the problem of solving the heat equation in (2.8) subject to the conditions in (2.9) and (2.10). Once K is found, u is determined from (2.13).

For later applications, we will now derive an asymptotic version of (2.13). In the present high-frequency diffraction problem, we are interested in the asymptotic solution of u as $k\to\infty$ or, more specifically, as

$$|k|\to\infty, \quad 0<\arg k<\pi/2. \tag{2.14}$$

It is well known (Sec. 4.1 of Ref. 13) that the behavior of u defined in (2.13) as $k\to\infty$ is governed by that of K as $t\to 0$. In all the later applications, it turns out that the leading term of K as $t\to 0$ assumes a special form, namely,

$$K(x,z,t)\sim\Lambda(x,z)(4\pi t)^{-1}\times\exp\{(-1/4t)[(x^2+z^2)^{1/2}+l]^2\}, \quad t\to 0, \tag{2.15}$$

where $\Lambda(x,z)$ and constant l are known. Inserting (2.15) into (2.13) and using the identity (formula 8.421(8) of Ref. 14)

$$\int_0^\infty (4\pi t)^{-1}\exp\{-(1/4t)[(x^2+z^2)^{1/2}+l]^2+k^2t\}dt = (i/4)H_0^{(1)}(k(x^2+z^2)^{1/2}+kl), \quad \pi/4<\arg k<3\pi/4$$

one obtains

$$u(x,z)\sim\Lambda(x,z)[\exp(ikl)]\left[\frac{(x^2+z^2)^{1/2}}{(x^2+z^2)^{1/2}+l}\right]^{1/2}u^i(x,z), \tag{2.16}$$

which is valid for

$$|k|\to\infty, \quad \pi/4<\arg k<3\pi/4. \tag{2.17}$$

A study of (2.16) reveals that (2.17) may be replaced by (2.14) by an analytical continuation argument, and without violating the radiation condition in (2.4). Summarizing, once K is determined and has the special form of (2.15), the desired solution u is given in (2.16) valid in the range (2.14).

3. SOLUTION BY PATH INTEGRAL

Using the path integral technique, $K(x,z,t)$ will be solved from (2.8)–(2.10) in two steps. First, we ignore the boundary condition (2.9), which is equivalent to the removal of the half-plane Σ in Fig. 1. The solution obtained in this simplified case is next modified to yield the solution in the original problem where Σ is present.

Ignoring the boundary condition on K, (2.8) and (2.10) have an exact solution

$$K_0(B;Q)=(4\pi t)^{-1}\exp[(-1/4t)(a+b)^2]. \tag{3.1}$$

We have alternatively written $K_0(x,z,t)$ evaluated at B as $K_0(B;Q)$ to emphasize the observation point B and the source point Q. The subscript 0 of K_0 signifies the absence of Σ. The solution in (3.1) may be expressed

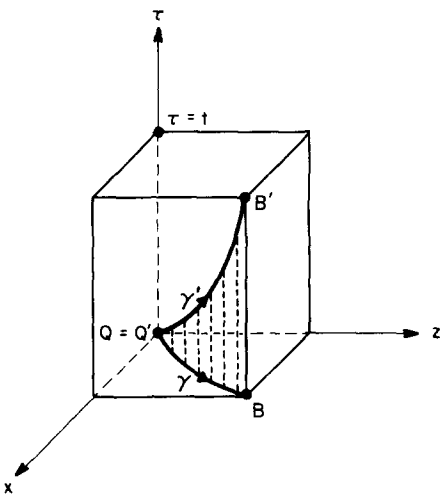


FIG. 2. Path γ' is a three-dimensional space curve from Q' to B'. Its projection in (x, z)-plane is γ.

in terms of a path integral over the Wiener measure. Let $x(\tau)$ and $z(\tau)$ be two single-valued continuous functions of "time" τ with end conditions

$$\begin{aligned} &\tau=0, \quad (x,z)=(0,0), \\ &\tau=t, \quad (x,z)=[(a+b)\sin\theta,\ (a+b)\cos\theta]. \end{aligned} \tag{3.2}$$

Then $[x(\tau),\ z(\tau),\ \tau]$ represents a space curve γ' going from $Q'=(Q,\tau=0)$ to $B'=(B,\tau=t)$ in three-dimensional (x, z, τ)-space,[15] as shown in Fig. 2. The projection of γ' on the (x, z)-plane is a planar curve γ. The set of all possible three-dimensional paths from Q' to B' is denoted by Γ. For a path γ', a functional $S[\gamma', t]$ is defined in the following manner. Let the interval $(0, t)$ be divided into N equal parts, each of length (t/N). Denote (x, z) at $\tau=nt/N$ by (x_n, z_n), for $n=0, 1, 2, \ldots, N$. In particular (x_0, z_0) coincide with the coordinates of Q, and (x_N, z_N) with those of B. As $N\to\infty$, the polygonal curve formed by connecting (x_{n-1}, z_{n-1}) to (x_n, z_n) for $n=1, 2, \ldots, N$ is an approximation of path γ'. Then S along γ' is defined by

$$S[\gamma',t]=\lim_{N\to\infty}\sum_{n=1}^{N}\frac{N}{4t}[(x_n-x_{n-1})^2+(z_n-z_{n-1})^2]. \tag{3.3}$$

The solution K_0 in (3.1) can be rewritten as a functional integral over all paths in Γ, namely,

$$\begin{aligned} &K_0(B;Q) \\ &=\lim_{N\to\infty}\left(\frac{N}{4\pi t}\right)^N\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}dx_1dx_2\cdots dx_{N-1}dz_1\cdots dz_{N-1} \\ &\times\exp\left\{\frac{-N}{4t}\sum_{n=1}^{N}[(x_n-x_{n-1})^2+(z_n-z_{n-1})^2]\right\}. \end{aligned} \tag{3.4}$$

The fact that (3.4) is identical to (3.1) may be verified by repeatedly using the integration identity

$$\begin{aligned} &\left(\frac{N}{4\pi t}\right)^{1/2}\left(\frac{N}{4\pi tm}\right)^{1/2}\int_{-\infty}^{\infty}dx\exp\left\{\frac{-N}{4t}\left[(x_a-x)^2+\frac{1}{m}(x-x_b)^2\right]\right\} \\ &=\left(\frac{N}{4\pi t(m+1)}\right)^{1/2}\exp\left[\frac{-N}{4t(m+1)}(x_a-x_b)^2\right], \quad \text{for } m>0. \end{aligned} \tag{3.5}$$

In fact, the expression behind the limit symbol in (3.4) is equal to $K_0(B, Q)$ for any N. Hence (3.4) is the limit of a sequence of identical terms. Alternatively, (3.3) and (3.4) are written as

$$S[\gamma',t]=\frac{1}{4}\int_0^t\left[\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dz}{d\tau}\right)^2\right]d\tau, \tag{3.6}$$

$$K_0(B,0)=\int_\Gamma \exp\{-S[\gamma',t]\}D\gamma'. \tag{3.7}$$

Several remarks about (3.6) and (3.7) are in order:

(i) Under the assumption that $x(\tau)$ and $z(\tau)$ are differentiable, (3.6) arises as the limit of the Riemann sum in (3.3).

(ii) (3.7) is a symbolic representation of (3.4), introduced first by Feynman. The integral is over all paths from Q' to B' in the three-dimensional (x, z, τ)-space.

(iii) Note that S in (3.6) is a functional of the three-dimensional path γ', which depends on both the planar path γ and its parameter τ. Different parametrizations of the same path γ may yield different values of S. More specifically, when replacing τ by a new parameter $\tau_1=f(\tau)$ subject to $f(0)=0$ and $f(t)=t$, it is easily seen that S evaluated from (3.6) generally changes.

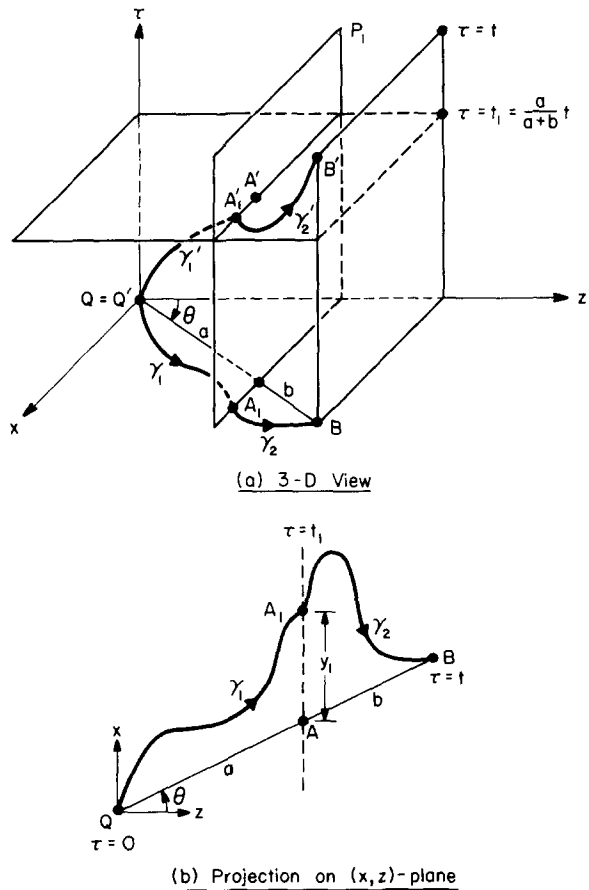


FIG. 3. All paths from Q' to B' may be grouped into two subsets. One contains all paths from Q' to a point A_1' on the infinite line $A'A_1'$, and the other contains all paths from A_1' to B'.

(iv) If one compares the present results with the path integral used in solving the Schrödinger equation, e.g., Eq. (1.5) and (1.6) in Ref. 1, S in (3.6) may be identified with the "action" of a free particle with mass $m=\frac{1}{2}$, and the K_0 in (3.7) with the "propagator" with the Planck's constant $h=-1$. We will adopt these two names for S and K_0.

The path integral in (3.4) may be put into a different form. Consider two sets of paths: Γ_1 contains all paths from $Q'=(Q,\ \tau=0)$ to $A_1'=(A_1,t_1)$ such as γ_1' in Fig. 3, and Γ_2 contains all paths from A_1' to $B'=(B,t)$ such as path γ_2'. Point A_1', which is a point on the infinite strip P_1, has coordinates $(x=a\sin\theta+y_1,\ z=a\cos\theta,\ \tau=t_1)$. Following (3.7) and (3.1), we find the propagators from Q' to A_1' and from A_1' to B' to be

$$K_0(A_1;Q)=\int_{\Gamma_1}\{\exp(-S[\gamma',t_1])\}D\gamma' \tag{3.8a}$$

$$=(4\pi t_1)^{-1}\exp\{(-1/4t_1)[(a\sin\theta+y_1)^2+(a\cos\theta)^2]\}, \tag{3.8b}$$

$$K_0(B;A_1)=\int_{\Gamma_2}\{\exp(-S[\gamma',t_2])\}D\gamma' \tag{3.8c}$$

$$=(4\pi t_2)^{-1}\exp\{(-1/4t_2)[(b\sin\theta-y_1)^2+(b\cos\theta)^2]\}. \tag{3.8d}$$

Note that the position of A_1' depends upon two parameters: y_1 and t_1 whose ranges are, respectively,

$$-\infty<y_1<\infty \quad\text{and}\quad 0<t_1<t. \tag{3.9}$$

If y_1 and t_1 are allowed to take all possible values in (3.9), the union of Γ_1 and Γ_2 recovers Γ. Based on this argument, one should expect the following result:

$$K_0(B:Q)=(4\pi t)^{-1/2}\int_0^{\tau}dt_1\,(4\pi t_1)^{1/2}(4\pi t_2)^{1/2}g(t_1)\times\int_{-\infty}^{\infty}dy_1K_0(B;A_1)K_0(A_1;Q), \tag{3.10}$$

where $t_2=t-t_1$ and $g(t_1)=1$. The three factors $(4\pi\bar{t})^{1/2}$, $\bar{t}=t,t_1$, and t_2 in (3.10) are normalization factors similar to that used in Eq. (2.31) in Ref. 2, p. 37. We substitute (3.8b) and (3.8d) into (3.10), evaluate the integrals [similar to (3.5)], and expect that the exact solution in (3.1) should be recovered. It is found that this is so only if

$$g(t_1)=\delta\left(t_1-\frac{a}{a+b}\,t\right), \tag{3.11}$$

where δ is a Dirac-delta function. Then (3.10) becomes

$$K_0(B;Q)=(4\pi t)^{-1/2}(4\pi t_1)^{1/2}(4\pi t_2)^{1/2}\times\int_{-\infty}^{\infty}dy_1K_0(B;A_1)K_0(A_1;Q), \tag{3.12a}$$

subject to the condition that

$$\frac{a}{t_1}=\frac{b}{t_2}=\frac{a+b}{t}. \tag{3.12b}$$

The implication of (3.12) is as follows. In evaluating the path integral in (3.4), it is permissible to group the paths into two sets: one including all paths from Q to an arbitrary point on the infinite line AA_1 (Fig. 2b) and the other including all paths from that point to B, provided that all points on line AA_1 are reached simultaneously at $\tau=t_1$ given in (3.12b). Referring to the three-dimensional space in Fig. 2a, we note that A_1' must be on the infinite line $A'A_1'$, instead of the infinite strip P_1 defined by (3.9).

Return to the original problem where Σ is present. As indicated in Ref. 1, the method of path integral can be again used provided that the space (x,z) is considered to be made of two sheets joining at Σ. Such a formulation should lead to a rigorous solution to the problem posed. However, in this paper, we adopt a different approach which appears simpler and yields only the dominant asymptotic solution of u.

To calculate the total field at B (Fig. 1) to the order of k^0 by conventional techniques, it is well known that the problem can be simplified by treating Σ as an "absorbing" screen (an opaque screen, or a screen whose reflection coefficient is zero). In the path integral formulation, we postulate that this simplification is equivalent to neglecting paths intersecting Σ such that the asymptotic solution derived from (2.8)—(2.10) is given by

$$K(B;Q)\sim(4\pi t)^{-1/2}(4\pi t_1)^{1/2}(4\pi t_2)^{1/2}\times\int_0^{\infty}dy_1K_0(B;A_1)K_0(A_1;Q),\quad t\to 0, \tag{3.13}$$

where t_1 and t_2 again satisfy (3.12b). Compare K in (3.13) with K_0 in (3.12), the exact solution when Σ is absent. We note that the integration limit of y_1 is from 0 to ∞ in (3.13) instead of from $-\infty$ to ∞. Using (3.8b), (3.8d), and (3.12b) in (3.13), one obtains

$$K(B;Q)\sim\tfrac{1}{2}(4\pi t)^{-1}\exp\{(-1/4t)(a+b)^2\},\quad t\to 0. \tag{3.14}$$

A comparison of (3.14) and (2.15) leads to the result that $\Lambda=\frac{1}{2}$ and $l=0$. Then the total field at point B for the diffraction problem sketched in Fig. 1 is determined from (2.16), namely,

$$u(b)\sim\tfrac{1}{2}u^i(B),\quad k\to\infty, \tag{3.15}$$

which agrees with the dominant term in the rigorous asymptotic solution given in (2.5). This simple result in (3.15) is expected as the presence of Σ "blocks" half of the paths from Q to B. The contribution of the present section, however, lies in the deduction of (3.12b) and (3.13) which will be used later in more complex diffraction problems.

4. TRANSMISSION THROUGH TWO STAGGERED HALF-PLANES

As an application of the path integral developed in the above two sections, consider the transmission of an electromagnetic wave through two staggered conducting half-planes Σ_1 and Σ_2 shown in Fig. 4$(\theta\neq 0)$. For a given incident field u^i in (2.7) from a line source at Q, the asymptotic solution of the total field u at C is to be determined. The point C, located on the line connecting Q, A, and B, is exactly on the incident shadow boundary.

Let us denote an arbitrary point in the aperture of Σ_1 by A_1 with coordinates $(x=a\sin\theta+y_1,\ z=a\cos\theta)$, and that in the aperture of Σ_2 by B_2 with coordinates $[x=(a+b)\sin\theta+y_2,\ z=(a+b)\cos\theta]$ (Fig. 3). All of the paths from Q at $\tau=0$ to C at $\tau=t$ are grouped into three sets[16]:

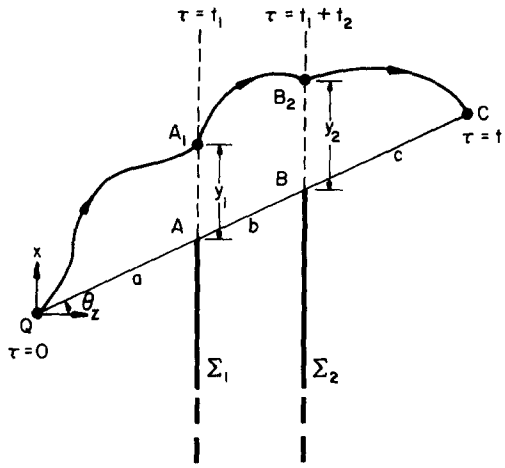


FIG. 4. Two staggered half planes Σ_1 and Σ_2 illuminated by the incident field from a line source at Q.

Set Γ_1: all paths from Q at $\tau=0$ to A_1 at $\tau=t_1$,

Set Γ_2: all paths from A_1 at $\tau=t_1$ to B_2 at $\tau=t_1+t_2$,

Set Γ_3: all paths from B_2 at $\tau=t_1+t_2$

to C at $\tau=t=t_1+t_2+t_3$.

As in (3.1), the propagators calculated from Γ_1, Γ_2, and Γ_3 are

$$K_0(A_1;Q)=(4\pi t_1)^{-1}\exp\{(-1/4t_1)[(a\sin\theta+y_1)^2+(a\cos\theta)^2]\}, \tag{4.1a}$$

$$K_0(B_2;A_1) \tag{4.1b}$$
$$=(4\pi t_2)^{-1}\exp\{(-1/4t_2)[(b\sin\theta+y_2-y_1)^2+(b\cos\theta)^2]\},$$

$$K_0(C;B_2) \tag{4.1c}$$
$$=(4\pi t_3)^{-1}\exp\{(-1/4t_3)[(c\sin\theta-y_2)^2+(c\cos\theta)^2]\}.$$

As in (3.13), we postulate that the asymptotic solution of the propagator from Q to C is given by

$$K(C;Q)\sim(4\pi t)^{-1/2}\left[\prod_{n=1}^{3}(4\pi t_n)^{1/2}\right]\int_0^\infty dy_1\int_0^\infty dy_2 \times K_0(C;B_2)K_0(B_2;A_1)K_0(A_1;0),\quad t\to 0. \tag{4.2}$$

The condition corresponding to (3.12b) is

$$\frac{a}{t_1}=\frac{b}{t_2}=\frac{c}{t_3}=\frac{a+b+c}{t}. \tag{4.3}$$

Substituting (4.3) and (4.1) into (4.2), one obtains

$$K(C;Q)\sim(4\pi t)^{-2}(a+b+c)^{3/2}(abc)^{-1/2}\exp\left[\frac{-(a+b+c)^2}{4t}\right] \times\int_0^\infty dy_1\int_0^\infty dy_2\exp\left\{\frac{-(a+b+c)}{4t} \times\left[\frac{1}{a}y_1^2+\frac{1}{b}(y_2-y_1)^2+\frac{1}{c}y_2^2\right]\right\}. \tag{4.4}$$

The double integral in (4.4) can be exactly evaluated (Eq. (II.54) in Ref. 8). The result is

$$K(C;Q)\sim\left\{\frac{1}{4}+\frac{1}{2\pi}\tan^{-1}\left[\frac{ac}{b(a+b+c)}\right]^{1/2}\right\} \times(4\pi t)^{-1}\exp\left[\frac{-1}{4t}(a+b+c)^2\right],\quad t\to 0. \tag{4.5}$$

Then it follows from (2.15) and (2.16) that the total field at C for the diffraction problem sketched in Fig. 4 is

$$u(C)\sim u^i(C)\left\{\frac{1}{4}+\frac{1}{2\pi}\tan^{-1}\left[\frac{ac}{b(a+b+c)}\right]^{1/2}\right\},\quad k\to\infty. \tag{4.6}$$

This solution agrees with the dominant term in the asymptotic solution derived by the uniform asymptotic theory[6,8] and for the special case $a\to\infty$ agrees with that derived by rigorous analytical methods.[5,7]

Consider the diffraction problem sketched in Fig. 5, whose configuration is obtained by moving Σ_2 in Fig. 4 to its complementary position. Using the path integral technique, it is readily seen that the propagator $K(C;Q)$ is again given in (4.4) except that y_2 is replaced by $-y_2$. This leads to a solution for the total field at C in Fig. 5, namely,

$$u(C)\sim u^i(C)\left\{\frac{1}{4}-\frac{1}{2\pi}\tan^{-1}\left[\frac{ac}{b(a+b+c)}\right]^{1/2}\right\},\quad k\to\infty, \tag{4.7}$$

which differs from (4.6) only in the sign of the arctangent term. As expected, note that $u(C)$ in (4.7) reduces to zero as $b\to 0$.

5. TRANSMISSION THROUGH TWO UNSTAGGERED HALF PLANES

The field solution given in (4.6) is asymptotic to the order k^0 as $k\to\infty$. When half-planes are unstaggered, (4.6) becomes imcomplete to that order. As shown in Ref. 8 by ray-optic techniques, the field contribution from the interaction between two half-planes is of order k^{-1} in the staggered case, and is increased to order k^0 in the unstaggered case. Thus, additional terms of order k^0 must be added to (4.6) in the unstaggered case. Those, additional terms will be calculated by the path integral technique in this section.

Let us reconsider the staggered case in Fig. 4. Among all paths from Q to C, only those in the neighborhood of path $QABC$ contribute significantly to the asymptotic solution of propagator $K(C;Q)$. This is evident from the fact that as $t\to 0$ the main contribution of the integral in (4.4) comes from the neighborhood of $y_1=0$ and $y_2=0$. Path $QABC$, of course, is precisely the incident ray in the geometrical theory of diffraction.[11,12] The fact that there is no other ray contributing to the field of order k^0 at C justifies our approximation

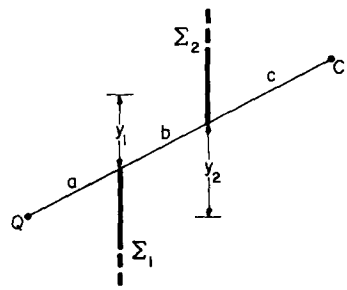


FIG. 5. Same diffraction problem as that in Fig. 3 except for the position of Σ_2.

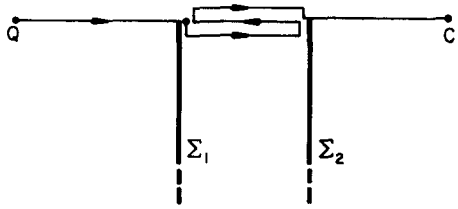


FIG. 6. A ray from Q to C after two specular reflections at Σ_1 and Σ_2.

in Sec. 4 that paths intersecting Σ_1 or Σ_2 can be ignored.[17]

When the plates are unstaggered, there exist additional rays which give rise to field contributions of order k^0 at C. One of these rays is shown in Fig. 6. This ray emanating from the source at Q reaches C after two specular reflections at Σ_1 and Σ_2. Since a specularly reflected field is of the same order as the incident one, the field at C on this ray is also or order k^0, the same as that on the direct ray from Q to C. Corresponding to the ray in Fig. 5, one has to consider path $QA_1B_2A_3B_4C$ in Fig. 7 for the evaluation of $K(C;Q)$ in the present unstaggered case. Let us assume that points A_1, B_2, A_3, B_4, and C are reached at $\tau=t_1$, t_1+t_2, $t_1+t_2+t_3$, $t_1+\cdots+t_4$, and $t_1+\cdots+t_5$, respectively. Following (3.12b) or (4.3), the following relation holds:

$$\frac{a}{t_1}=\frac{b}{t_2}=\frac{b}{t_3}=\frac{b}{t_4}=\frac{c}{t_5}=\frac{a+3b+c}{t}. \tag{5.1}$$

The propagators associated with those points are

$$K_0(A_1;Q) = (4\pi t_1)^{-1}\exp[(-1/4t_1)(a^2+y_1^2)], \tag{5.2a}$$

$$K_0(B_2;A_1) = (4\pi t_2)^{-1}\exp\{(-1/4t_2)[b^2+(y_1+y_2)^2]\}, \tag{5.2b}$$

$$K_0(A_3;B_2) = (4\pi t_3)^{-1}\exp\{(-1/4t_3)[b^2+(y_2-y_3)^2]\}, \tag{5.2c}$$

$$K_0(B_4;A_3) = (4\pi t_4)^{-1}\exp\{(-1/4t_4)[b^2+(y_3+y_4)^2]\}, \tag{5.2d}$$

$$K_0(C;B_4) = (4\pi t_5)^{-1}\exp\{(-1/4t_5)(c^2+y_4^2)\}. \tag{5.2e}$$

As in (3.13) or (4.2), the asymptotic solution of the propagator from Q to C arising from paths as the one shown in Fig. 7 is given by

$$K^{(2)}(C;Q)\sim(4\pi t)^{-1/2}\left[\prod_{n=1}^{5}(4\pi t_n)^{1/2}\right]\int_0^\infty dy_1\cdots\int_0^\infty dy_4 \times K_0(C;B_4)K_0(B_4;A_3)\cdots K_0(A_1;Q), \quad t\to 0. \tag{5.3}$$

The superscript of $K^{(2)}$ signifies the contribution to K from those paths which are twice reflected from Σ_1 and Σ_2. Substituting (5.1) and (5.2) into (5.3), one obtains

$$K^{(2p)}(C;Q) \sim (4\pi t)^{-(p+2)}(abc)^{-1/2}b^{-p}d^{(2p+3)/2}\left[\exp\frac{-d^2}{(4t)}\right] \times\int_0^\infty dy_1\int_0^\infty dy_2\cdots\int_0^\infty dy_{2p+2} \times\exp\left\{\frac{-d}{4t}\left[\frac{1}{a}y_1^2+\frac{1}{b}(y_1+y_2)^2+\frac{1}{b}\sum_{n=2}^{2p}(y_n-y_{n+1})^2 +\frac{1}{b}(y_{2p+1}+y_{2p+2})^2+\frac{1}{c}y_{2p+2}\right]\right\}, \quad t\to 0, \tag{5.4}$$

where $d=a+(2p+1)b+c$ and $p=1$. The integral in (5.4) can be simplified by the following manipulations. First, we perform a change of variables by letting

$$\left(\frac{d}{4bt}\right)^{1/2}y_n\to y_n, \quad n=1,2,\ldots,2p+2. \tag{5.5}$$

Second, we simplify the integration with respect to y_{2p+2} with the help of the identities

$$\int_0^\infty F(\alpha)d\alpha=\int_{-\infty}^{\infty}F(-\alpha)d\alpha-\int_0^\infty F(-\alpha)d\alpha, \tag{5.6a}$$

$$\int_{-\infty}^{\infty}d\alpha\exp\left\{-\left[\left(\frac{c}{b+c}\right)^{1/2}\beta-\left(\frac{b+c}{c}\right)^{1/2}\alpha\right]^2\right\}=\left(\frac{\pi c}{b+c}\right)^{1/2}, \tag{5.6b}$$

where $\alpha=y_{2p+2}$ and $\beta=y_{2p+1}$. After these manipulations, (5.4) becomes

$$K^{(2p)}(C;Q) \sim\left(\frac{bd}{ac}\right)^{1/2}\left[\left(\frac{c}{b+c}\right)^{1/2}J_{2p+1}\left(\frac{b}{a},\frac{b}{b+c}\right) -J_{2p+2}\left(\frac{b}{a},\frac{b}{c}\right)\right](4\pi t)^{-1}\exp(-d^2/4t), \quad t\to 0, \tag{5.7}$$

where

$$J_n(\alpha,\beta)=(\pi)^{-n/2}\int_0^\infty dy_1\int_0^\infty dy_2\cdots\int_0^\infty dy_n \times\exp\{-[\alpha y_1^2+(y_1+y_2)^2+\sum_{m=2}^{n-1}(y_m-y_{m+1})^2+\beta y_n^2]\}. \tag{5.8}$$

With $p=1$, (5.7) gives the final asymptotic solution of $K^{(2)}$ contributed from the twice reflected paths.

In addition to the twice reflected paths, there are paths going from Q to C after $2p$ reflections between

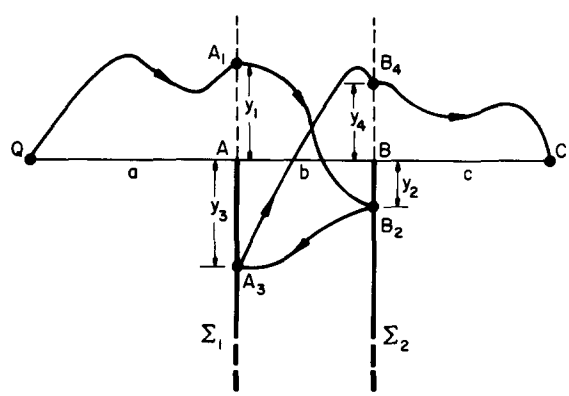


FIG. 7. Two unstaggered half planes Σ_1 and Σ_2 illuminated by the incident field from a line source at Q.

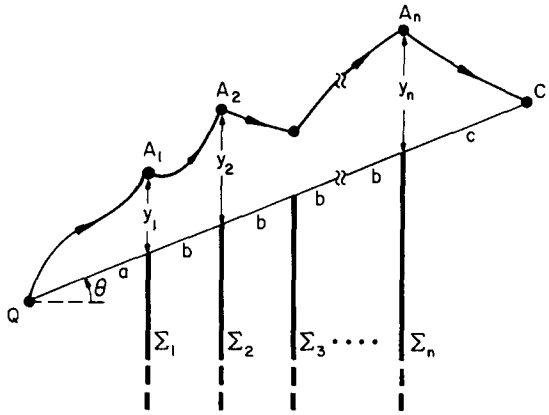


FIG. 8. n equally spaced staggered half planes illuminated by the incident field from a line source at Q.

Σ_1 and Σ_2. Their contributions are precisely $K^{(2p)}$ given in (5.7). Then, the propagator $K(Q;C)$ is the superposition of all $\{K^{(2p)}\}$ with $p=0,1,2,\ldots$, namely,

$$K(C;Q)\sim(4\pi t)^{-1}\sum_{p=0}^{\infty}\left(\frac{bd}{ac}\right)^{1/2}\left[\left(\frac{c}{b+c}\right)^{1/2}J_{2p+1}\left(\frac{b}{a},\frac{b}{b+c}\right)-J_{2p+2}\left(\frac{b}{a},\frac{b}{c}\right)\right]\exp(-d^2/4t),\quad t\to 0. \tag{5.9}$$

A comparison of (5.9) and (2.15) determines Λ and l. The asymptotic field solution at C for the diffraction problem sketched in Fig. 5 is found from (2.16), namely,

$$u(C)\sim u^i(C)\left\{\frac{1}{4}+\frac{1}{2\pi}\tan^{-1}\left[\frac{ac}{b(a+b+c)}\right]^{1/2}+\sum_{p=1}^{\infty}(\exp i2pkb)\left[\frac{b(a+b+c)}{ac}\right]^{1/2}\times\left[\left(\frac{c}{b+c}\right)^{1/2}J_{2p+1}\left(\frac{b}{a},\frac{b}{b+c}\right)-J_{2p+2}\left(\frac{b}{a},\frac{b}{c}\right)\right]\right\},\quad k\to\infty. \tag{5.10}$$

For the special case when the source is at an infinitely large distance away from Σ_1, i.e., $a\to\infty$, (5.10) is identical to the dominant term of order k^0 in Eq. (II.62) of Ref. 8 obtained by the uniform asymptotic theory. [Note the equivalent notations in (5.10) and (II.62): $b\to a$, $c\to r_0$, and $J_n(0,b/c)\to J_{n,0}(c)$.] Furthermore, if both the source and the observation points are far away, i.e., (b/a) and $(b/c)\to 0$, the integral $J_n(0,0)$ in (5.8) can be explicitly evaluated by several methods: one involving transformation of variables in the n-dimensional space and generating-function techniques, one using the Wiener–Hopf technique, and one based on probabilistic interpretation (random walk) of the integrals. Details of this evaluation are given in Ref. 18, while only the relevant final result is quoted below:

$$J_n(0,0)=1/2\pi\sqrt{n-1},\quad n=2,3,\ldots. \tag{5.11}$$

Using (5.11) in (5.10), one obtains the asymptotic field solution at C when both Q and C (Fig. 6) are far away from Σ_1 and Σ_2:

$$u(C)\sim u^i(C)\left\{\frac{1}{4}+\frac{1}{2\pi}\tan^{-1}\left(\frac{ac}{b(a+c)}\right)^{1/2}+\frac{1}{2\pi}\left(\frac{b(a+c)}{ac}\right)^{1/2}\times\sum_{p=1}^{\infty}[(2p)^{-1/2}-(2p+1)^{-1/2}]\exp i2pkb\right\},\quad k\to\infty,\ \frac{b}{a}\to 0,\ \frac{b}{c}\to 0. \tag{5.12}$$

For the special case that $(a/c)\gg 1$, (5.12) recovers the asymptotic expansion of the exact solution obtained by the Wiener–Hopf technique (see Eq. (II.68) or (A6) in Ref. 8).

6. TRANSMISSION THROUGH AN ARRAY OF STAGGERED PARALLEL HALF-PLANES

As a final example, consider the transmission through n equally spaced staggered parallel half-planes as shown in Fig. 8. For an incident field u^i from the line source at Q, the asymptotic solution of the total field at C is to be determined, where C falls exactly on the incident shadow boundary. Steps for solving this problem are very similar to those used in connection with Figs. 4 and 7. The final solution is found to be

$$u(C)\sim u^i(C)L_n(b/a,\,b/c),\quad k\to\infty, \tag{6.1a}$$

where

$$L_n(\alpha,\beta)=[\alpha+\beta+(n-1)\alpha\beta]^{1/2}\pi^{-n/2}\int_0^\infty dy_1\int_0^\infty dy_2\cdots\int_0^\infty dy_n\times\exp\{-[\alpha y_1^2+\sum_{m=1}^{n-1}(y_m-y_{m+1})^2+\beta y_n^2]\}. \tag{6.1b}$$

The solution in (6.1) is valid for $n=1,2,3,\ldots$. The symmetry in a and c in (6.1) confirms the reciprocity principle of the electromagnetic field, namely, the field at C due to a source at Q is identical to the field at Q due to the same source at C. The integral L_n has been studied by J. Boersma in an unpublished note, the essence of which is summarized in the Appendix. We quote below some of his final results:

$$L_1(\alpha,\beta)=\tfrac{1}{2}, \tag{6.2}$$

$$L_2(\alpha,\beta)=\tfrac{1}{4}+(1/2\pi)\tan^{-1}(1/\sqrt{\alpha+\beta+\alpha\beta}\,), \tag{6.3}$$

$$L_3(\alpha,\beta)=\tfrac{1}{8}+(1/4\pi)[\tan^{-1}(1/\sqrt{2}\gamma)+\tan^{-1}(\sqrt{\alpha+1}/\gamma)+\tan^{-1}(\sqrt{\beta+1}/\gamma)], \tag{6.4}$$

where $\gamma=\sqrt{\alpha+\beta+2\alpha\beta}$. The use of (6.2) and (6.3) in (6.1a) recovers the known results in (3.15) and (4.6). The use of (6.4) in (6.1a) gives a new result for diffraction by three staggered parallel half planes ($n=3$ in Fig. 8), namely,

$$u(C)\sim u^i(C)\left\{\frac{1}{8}+\frac{1}{4\pi}\tan^{-1}\left[\frac{ac}{2b(a+2b+c)}\right]^{1/2}+\frac{1}{4\pi}\tan^{-1}\left[\frac{(a+b)c}{b(a+2b+c)}\right]^{1/2}+\frac{1}{4\pi}\tan^{-1}\left[\frac{a(b+c)}{b(a+2b+c)}\right]^{1/2}\right\}. \tag{6.5}$$

For a general n, Boersma has shown that

$$L_n(\alpha,\beta)=\tfrac{1}{2}\pi^{-n/2}\Gamma(\tfrac{1}{2}n)S_n,$$

where S_n is the surface area of the spherical n-gon that is cut out in the n-dimensional unit sphere by G_n, and G_n is a certain polyhedral cone where it is bounded by

n hyperplanes through the center of the sphere. It appears that S_n cannot be determined in a simple closed form. For some special values of α and β, L_n in (6.1b) has been explicitly determined as given in (A9)—(A13) in the Appendix. Using those results, we obtain the following field solutions which hold for any n for the diffraction problem in Fig. 8.

(i) If $a=\infty$ and $c=\infty$ (far-away source and observation point),

$$u(C)\sim\tfrac{1}{2}u^i(C). \tag{6.6}$$

(ii) If $a=\infty$ and $c=0$ (far-away source, and observation point at the edge of the last half plane), or $a=0$ and $c=\infty$,

$$u(C)\sim\frac{\Gamma(n-\frac{1}{2})}{2(n-1)!\,\Gamma(\frac{1}{2})}u^i(C). \tag{6.7}$$

(iii) If $a=b$ and $c=0$, or $a=0$ and $c=b$,

$$u(C)\sim\frac{1}{2n}u^i(C). \tag{6.8}$$

(iv) If $a=0$ and $c=0$ (source and observation point at the edges of the first and the last half planes),

$$u(C)\sim\frac{1}{4(n-1)}u^i(C). \tag{6.9}$$

(v) If $a=b=c$,

$$u(C)\sim\frac{1}{n+1}u^i(C). \tag{6.10}$$

(vi) If $a=\infty$ and $c=b$, or $a=b$ and $c=\infty$,

$$u(C)\sim\frac{\Gamma(n+\frac{1}{2})}{n!\,\Gamma(\frac{1}{2})}u^i(C). \tag{6.11}$$

All the results given in (6.5)—(6.11) are new results for the diffraction problem under consideration. It appears that these results cannot be simply obtained by other methods known to us.

7. CONCLUSION

In the present application of path integrals to the edge diffraction problems, the key is the postulate in (3.13). Loosely speaking, this postulate states that in determining the dominant asymptotic solution of propagator $K(B;Q)$ in Fig. 1, the presence of the obstacle Σ is accounted for by ignoring the contributions from the paths that intersect Σ. Based on this postulate, we have been able to derive known results in (4.6) and (5.12) in a much simpler fashion than those reported in the literature, and to derive new results in (4.7), (5.10), and those presented in Sec. 6.

Our success with path integrals so far is restricted to the following two cases: (i) The observation point for the field is located exactly on the incident shadow boundary, and (ii) only the leading term of order k^0 in the asymptotic field solution has been determined. Further work must be done in order to relax these two restrictions.

ACKNOWLEDGMENTS

From Professor G. A. Deschamps in 1970 the author learned of the possible application of path integrals to electromagnetics. Using a different approach (G. A. Deschamps and S. W. Lee, "Feyneman path integral and electromagnetic problems," paper presented in URSI Meeting, Amherst, Mass., October 1976), Deschamps first obtained the results in (4.6). The author wishes also to express his gratitude to Professor J. Boersma who read the original manuscript and suggested many improvements.

APPENDIX: EVALUATION OF INTEGRAL $L_n(\alpha,\beta)$

The integral $L_n(\alpha,\beta)$ is defined in (6.1b). Its evaluation is an interesting mathematical problem, which has been studied by J. Boersma and is presented in an unpublished note. Some of his results are summarized below.

(i) The evaluation of $L_1(\alpha,\beta)$ is trivial, and that of $L_2(\alpha,\beta)$ is very similar to that of $J_{2,0}(r_0)$ in Eq. (II.54) of Ref. 8. The final results of L_1 and L_2 are given in (6.2) and (6.3).

(ii) To evaluate $L_3(\alpha,\beta)$, we introduce the new variables

$$y_1-\frac{y_2}{\alpha+1}=z_1(\alpha+1)^{-1/2}, \tag{A1a}$$

$$y_2=\left[\frac{(\alpha+1)(\beta+1)}{\alpha+\beta+2\alpha\beta}\right]^{1/2}z_2, \tag{A1b}$$

$$y_3-\frac{y_2}{\beta+1}=z_3(\beta+1)^{-1/2}. \tag{A1c}$$

Then $L_3(\alpha,\beta)$ is reduced to

$$L_3(\alpha,\beta)=\pi^{-3/2}\int\int\int_G dz_1dz_2dz_3\exp(-z_1^2-z_2^2-z_3^2), \tag{A2}$$

where the domain of integration is given by

$$G:\ z_1+pz_2\geq 0,\quad z_2\geq 0,\quad z_3+qz_2\geq 0, \tag{A3}$$

with

$$p=\left(\frac{\beta+1}{\alpha+\beta+2\alpha\beta}\right)^{1/2},\quad q=\left(\frac{\alpha+1}{\alpha+\beta+2\alpha\beta}\right)^{1/2}. \tag{A4}$$

If spherical coordinates are introduced, then the r-dependence part of the triple integral becomes

$$\int_0^\infty r^2\exp(-r^2)\,dr=\tfrac{1}{4}\pi^{1/2}. \tag{A5}$$

Thus, $L_3(\alpha,\beta)$ reduces to

$$L_3(\alpha,\beta)=\frac{1}{4\pi}S, \tag{A6}$$

where S is the surface area of the spherical triangle that is cut out in the unit sphere by the planes defined in (A3). From spherical trigonometry, it is found that

$$S=\pi/2+\tan^{-1}p+\tan^{-1}q+\tan^{-1}[pq(1+p^2+q^2)^{-1/2}]. \tag{A7}$$

Substitution of (A7) and (A4) into (A6) gives the final answer in (6.4).

(iii) The same method described in (ii) is used to evaluate $L_n(\alpha,\beta)$ for general n. The result is given in the equation after (6.5), where the surface area S_n has not been explicitly determined.

(iv) $L_n(\alpha,\beta)$ may be expressed in terms of $J_n(\alpha,\beta)$ defined in (5.8), namely,

$$L_n(\alpha,\beta)=\tfrac{1}{2}-[\alpha+\beta+(n-1)\alpha\beta]^{1/2}$$
$$\times\sum_{m=0}^{n-2}(1+m\alpha)^{-1/2}J_{n-m}\left(\frac{\alpha}{1+m\alpha},\ \beta\right). \tag{A8}$$

Setting $\alpha=\beta=0$, one finds

$$L_n(0,0)=\tfrac{1}{2}, \tag{A9}$$

which is used to derive the field solution in (6.6).

(v) $L_n(1,1)$ is identical to $(n+1)^{1/2}\,I_{n,0}(a)$ determined in Ref. 18, viz.,

$$L_n(1,1)=\frac{1}{n+1}, \tag{A10}$$

which is used to derive (6.10). $L_n(0,1)$ is identical to $I_{n,0}(\infty)$ determined in Ref. 18, namely,

$$L_n(0,1)=\Gamma(n+\tfrac{1}{2})/n!\,\Gamma(\tfrac{1}{2}), \tag{A11}$$

which is used to derive (6.11). Furthermore, from (6.1b) it can be shown that

$$L_n(\alpha,\infty)=\tfrac{1}{2}L_{n-1}(\alpha,1). \tag{A12}$$

The relations in (A.10)—(A.12) are used to derive (6.7) and (6.8). Finally, the result

$$L_n(\infty,\infty)=\tfrac{1}{4}L_{n-2}(1,1)=1/4(n-1) \tag{A13}$$

is used to derive (6.9).

[1]J.B. Keller and D.W. McLaughlin, "The Feynman integral," Amer. Math. Monthly **82**, 451—65 (1975).
[2]R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integral* (McGraw-Hill, New York, 1965).
[3]B. Noble, *Methods Based on the Wiener-Hopf Technique* (Pergamon, London, 1958).
[4]R. Mittra and S.W. Lee, *Analytical Techniques in the Theory of Guided Waves* (Macmillan, New York, 1971).
[5]D.S. Jones, "Double knife-edge diffraction and ray theory," Quart. J. Mech. Appl. Math. **26**, 1—18 (1973).
[6]J. Boersma, "Diffraction by two parallel half-planes," Quart J. Mech. Appl. Math. **28**, 405—25 (1975).
[7]Y. Rahmat-Samii and R. Mittra, "On the investigation of diffracted fields at the shadow boundaries of stagged parallel plates—a spectral domain approach," Radio Sci. **12**, 659—70 (1977).
[8]S.W. Lee and J. Boersma, "Ray-optical analysis of field on shadow boundaries of two parallel plates," J. Math. Phys. **16**, 1746—64 (1975).
[9]V.S. Buslaev, "Continuum integrals and the asymptotic behavior of the solutions of parabolic equations as $t\to 0$. Applications to diffraction," in *Topics in Mathematical Physics, Vol.* 2, edited by M. Sh. Birman (Consultants Bureau, English Transl., New York, 1968), pp. 67—86.
[10]M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1964), 2nd ed.
[11]R.M. Lewis and J. Boersma, "Uniform asymptotic theory of edge diffraction," J. Math. Phys. **10**, 2291—2305 (1969).
[12]J.B. Keller, "Geometrical theory of diffraction," J. Opt. Soc. Amer. **52**, 116—30 (1962).
[13]N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals* (Holt, Rinehart and Winston, New York, 1975).
[14]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), 4th ed.
[15]The path, of course, must satisfy causality with respect to time τ.
[16]From here on, we always work in the projected (x,z)-space (Fig. 2b) instead of the three-dimensional (x,z,τ)-space (Fig. 2a).
[17]Through private communication, J. Boersma has recently calculated the $K^{(2)}(C;Q)$ defined in (5.3) for the *staggered* case by the present path integral technique. His result shows that the twice-reflected paths indeed give an asymptotic contribution of order k^{-1} to the field $u(C)$.
[18]J. Boersma, "On certain multiple integrals occurring in a waveguide scattering problems," to appear in SIAM J. Math. Anal., 1978.

Cylindrically symmetric dust distributions in rigid rotation in Brans–Dicke theory

Nikhilendu Bandyopadhyay

Physics Department, Presidency College, Calcutta, India
(Received 18 May 1977)

A solution for the field equations of Brans–Dicke theory for the case of a cylindrically symmetric dust distribution in rigid rotation has been obtained. The exact solution shows the existence of closed timelike lines as also infinities of mass density, $B-D$ scalar, etc., at finite proper distance from the axis of symmetry just as in the corresponding general relativistic case.

1. INTRODUCTION

It is usually held that the Brans–Dicke (BD) theory is more consistent with Machian ideas than the general theory of relativity. Some support for this may be found in Banerjee's result[1] that there exists no analog to the Gödel solution in BD theory. However, there is another well-known solution due to Van Stockum[2] in general relativity where again the matter is in rotation. Here we address ourselves to the question whether a similar solution, i.e., a cylindrically symmetric (nonuniform) dust distribution in rigid rotation is consistent with the BD theory. The exact solutions to the field equations obtained below show the existence of closed timelike lines as in Van Stockum's solution. Here there is also a singularity corresponding to infinite density at a finite proper distance from the axis of symmetry.

2. THE FIELD EQUATIONS AND THEIR INTEGRATION

We consider a stationary cylindrically symmetric line element,

$$ds^2 = g_{00}\,dt^2 - \exp(2\psi)(dr^2+dz^2) - L\,d\Phi^2 + 2m\,d\Phi\,dt, \tag{2.1}$$

where all the metric tensor components are functions of r alone. In our discussion, the coordinates t, r, z, Φ will be numbered as $0, 1, 2, 3$, respectively. The BD field equations for a distribution of incoherent matter are:

$$R^\mu_\nu = \frac{8\pi}{\phi}\left(T^\mu_\nu - \tfrac{1}{2}T\,\delta^\mu_\nu\right) + \frac{\omega}{\phi^2}\,\phi^\mu\phi_\nu + \frac{\phi^\mu{}_{;\nu}}{\phi} + \tfrac{1}{2}\delta^\mu_\nu\frac{\Box\phi}{\phi}, \tag{2.2}$$

$$T^\mu_\nu = \rho v^\mu v_\nu, \tag{2.3}$$

$$\Box\phi = \frac{8\pi\rho}{3+2\omega}, \tag{2.4}$$

where ϕ is the BD scalar, other symbols having their usual meaning.

We assume the coordinate system to be moving so that $v^\mu = \delta^\mu_0\, g_{00}^{-1/2}$.

As there is no nongravitational interaction, the velocity vector is geodesic (as indeed follows from the divergence indentity involving the energy–momentum tensor):

$$v^\mu_{;\nu}v^\nu = 0,$$

and so $g_{00} =$ constant. Without loss of generality one can put this constant equal to unity,

$$g_{00} = 1. \tag{2.5}$$

Using the notation

$$D^2 = L + m^2, \tag{2.6}$$

the field equations can now be written explicitly as

$$\left(\frac{mm_1}{2D}\right)_1 = \frac{\sqrt{-g}}{\phi}\,4\pi\rho - \frac{mm_1}{2D}\,\frac{\phi_1}{\phi} + \tfrac{1}{2}\sqrt{-g}\,\frac{\Box\phi}{\phi}, \tag{2.7}$$

$$\left(\frac{L+mm_1}{2D}\right)_1 = -\frac{\sqrt{-g}}{\phi}\,4\pi\rho - \frac{L_1+mm_1}{2D}\,\frac{\phi_1}{\phi} + \tfrac{1}{2}\sqrt{-g}\,\frac{\Box\phi}{\phi}, \tag{2.8}$$

$$\left(\frac{mL_1 - Lm_1}{2D}\right)_1 = -\frac{\sqrt{-g}}{\phi}\,m8\pi\rho - \frac{mL_1 - Lm_1}{2D}\,\frac{\phi_1}{\phi}, \tag{2.9}$$

$$\left(\frac{m_1}{2D}\right)_1 = -\frac{m_1}{2D}\,\frac{\phi_1}{\phi}, \tag{2.10}$$

$$-D\psi_{11} - D\psi_1 = \frac{\sqrt{-g}}{\phi}\,4\pi\rho + D\psi_1\,\frac{\phi_1}{\phi} - \frac{1}{2}\,\frac{\Box\phi}{\phi}\sqrt{-g}, \tag{2.11}$$

$$-D\psi_{11} + \frac{m_1^2}{2D} + D_1\psi_1 - D_{11} = \frac{\sqrt{-g}}{\phi}\,4\pi\rho + \omega D\,\frac{\phi_1^2}{\phi^2} + D\,\frac{\phi_{11}}{\phi} - D\psi_1\,\frac{\phi_1}{\phi} - \tfrac{1}{2}\sqrt{-g}\,\frac{\Box\phi}{\phi}, \tag{2.12}$$

where the subcript 1 indicates differentiation with respect to r.

Equation (2.10) readily gives the first integral

$$\frac{m_1}{2D} = \frac{b}{\phi} \quad (b = \text{const}). \tag{2.13}$$

Again from Eqs. (2.7) and (2.8) one has

$$(D_1\phi)_1 = \sqrt{-g}\,\Box\phi, \tag{2.14}$$

whereas from the definition of $\Box\phi$ one has

$$\sqrt{-g}\,\Box\phi = -(D\phi_1)_1. \tag{2.15}$$

Combining Eqs. (2.14) and (2.15) one has the relation

$$D\phi = ar, \tag{2.16}$$

a being an arbitrary constant.

Using Eq. (2.13), Eq. (2.7) yields

 0022-2488/78/1906-1423$1.00

$$\frac{m_1^2}{2D}=\frac{\sqrt{-g}}{\phi}4\pi\rho+\tfrac{1}{2}\sqrt{-g}\,\frac{\Box\phi}{\phi}. \tag{2.17}$$

Equation (2.17), in conjunction with Eqs. (2.13), (2.15), and (2.16) gives

$$-\frac{2b^2r}{\phi}=(2+\omega)\left(r\frac{\phi_1}{\phi}\right)_1,$$

which reduces to

$$\xi_{xx}=-\frac{2b^2}{2+\omega}\exp[-2(\xi-x)], \tag{2.18}$$

with the substitutions

$$\phi\sim\exp(\xi),\quad r\sim\exp(x), \tag{2.19}$$

the subscript x indicating differentiation with respect to x. To integrate Eq. (2.18), let

$$y=\xi-x. \tag{2.20}$$

Then Eq. (2.18) reduces to

$$y_{xx}=-\frac{2b^2}{2+\omega}\exp(-2y),$$

with the first integral

$$y_x^2=\lambda+\frac{2b^2}{2+\omega}\exp(-2y), \tag{2.21}$$

λ being the constant of integration. With the substitutions

$$z^2=\lambda+\frac{2b^2}{2+\omega}\exp(-2y),\quad \lambda=\mu^2, \tag{2.22}$$

Eq. (2.21) integrates to

$$\frac{\mu+z}{\mu-z}=\exp[2\mu(x+\nu)], \tag{2.23}$$

where ν is the constant of integration. Returning to original variables, ϕ has the form

$$\phi=i\left(\frac{2}{2+\omega}\right)^{1/2}\frac{b}{\mu}r\cosh\{\mu\ln(r/\alpha)\}, \tag{2.24}$$

where $\nu=-\ln\alpha$.

In expression (2.24) one must choose $\mu=\pm1$ in order that ϕ remains finite at the origin. For definiteness we choose $\mu=1$. Again to make ϕ real we let $\alpha=-i\beta$, where β is real. Thus ϕ has the form

$$\phi=\frac{1}{2}\left(\frac{2}{2+\omega}\right)^{1/2}b\beta\left(1-\frac{r^2}{\beta^2}\right). \tag{2.25}$$

Normalizing ϕ to unity at the origin we have

$$\beta=\sqrt{2(2+\omega)}/b. \tag{2.26}$$

Obviously with $\omega\to\infty$, $\phi\to1$, corresponding to the general relativity condition.

ψ is now obtained from Eq. (2.11), which has the first integral

$$\psi_1=(1+\omega)\frac{\phi_1}{\phi}. \tag{2.27}$$

Here the constant of integration has been put equal to zero to avoid singularity in the metric at the origin. Equation (2.27) further integrates to

$$\psi=(1+\omega)\ln\left(1-\frac{r^2b^2}{2(2+\omega)}\right), \tag{2.28}$$

where the constant of integration has again been put equal to zero in accordance with the argument given by Som and Raychaudhuri.[3] Here also with

$$\omega\to\infty,\ \psi\to\lim_{\omega\to\infty}\ln\left(1-\frac{r^2b^2}{2(2+\omega)}\right)^{2+\omega}=-\tfrac{1}{2}r^2b^2.$$

To find m, one makes use of Eq. (2.13) and (2.16) to get

$$m_1=2abr\left(1-\frac{r^2b^2}{2(2+\omega)}\right)^{-2},$$

with the integral

$$m=\frac{2a(2+\omega)}{b}\left(1-\frac{r^2b^2}{2(2+\omega)}\right)^{-1}+\text{const.} \tag{2.29}$$

The constant is chosen such that $m\to0$ as $r\to0$, i.e.,

$$\text{const}=-\frac{2a(2+\omega)}{b}.$$

Now, choosing $a=1$ to obtain correspondence with Van Stockum's results in the limit $\omega\to\infty$,

$$\lim_{\omega\to\infty}m=\lim_{\omega\to\infty}\frac{2a(2+\omega)}{b}\left[1+\frac{r^2b^2}{2(2+\omega)}+\left(\frac{r^2b^2}{2(2+\omega)}\right)^2+\cdots-1\right]$$
$$=abr^2,$$

m takes the form

$$m=br^2\left(1-\frac{r^2b^2}{2(2+\omega)}\right)^{-1}. \tag{2.30}$$

An expression for L is obtained from Eq. (2.16) as

$$L=\frac{(1-b^2r^2)r^2}{1-(r^2b^2)/2(2+\omega)}. \tag{2.31}$$

With $\omega\to\infty$, $L\to(1-b^2r^2)r^2$. Also from Eq. (2.4), ρ comes out as

$$8\pi\rho=\frac{2(3+2\omega)}{2+\omega}b^2\left(1-\frac{r^2b^2}{2(2+\omega)}\right)^{-(4+2\omega)}. \tag{2.32}$$

With $\omega\to\infty$,

$$8\pi\rho\to4b^2\lim_{\omega\to\infty}\left(1-\frac{r^2b^2}{2(2+\omega)}\right)^{-2(2+\omega)}=4b^2\exp(r^2b^2/2).$$

CONCLUDING REMARKS

As we have already noted, in the limit $\omega\to\infty$, the solution passes over to that of Van Stockum in general relativity. Also similar to that case, there are closed timelike lines for $r>1/b$ which have a singularity at $r=r'=[2(2+\omega)]^{1/2}/b$, where the density, the curvature scalar, and the BD scalar, etc., are infinite. This singularity evidently occurs at a finite proper distance from the axis of symmetry, since $\int_0^{r'}\exp(\psi)dr$ is convergent. A similar singularity in the general relativistic case has been pointed out by Shepley.[4] However, in the general relativistic case $\omega\to\infty$ and hence the radial coordinate r' of the singularity is infinite, notwithstanding the finiteness of the proper distance of the singularity from the axis of symmetry.

However, the present solution cannot be considered to be a representation of the real universe as it departs seriously from homogeneity and isotropy and on account of its stationary character cannot exhibit any isotropic red shift as is actually observed.

ACKNOWLEDGMENT

The author is thankful to Professor A. K. Raychaudhuri of the Physics Department, Presidency College, Calcutta, for suggesting the problem and extending helpful comments.

[1]S. Banerjee, Phys. Rev. D **9**, 877 (1974).
[2]W. J. Van Stockum, Proc. Roy. Soc. Edinburgh **57**, 135 (1937).
[3]M. M. Som and A. K. Raychaudhuri, Proc. Roy. Soc. A **304**, 81 (1968).
[4]L. C. Shepley, Proc. Nat. Acad. Sci. U.S.A. **52**, 1403 (1964).

A method for computing scattering phase shifts and eigenvalues of the Schrödinger equation with singular potentials

Friedrich F. Naundorf

Institut für Angewandte Mathematik der Universität Heidelberg, 69 Heidelberg, Germany
(Received 3 January 1977)

An algorithm is given for computing the scattering phase shifts and eigenvalues of the Schrödinger radial equation with singular potentials. As an application we treat the hydrogen atom with fine structure in the nonrelativistic theory.

1. INTRODUCTION

In the nonrelativistic theory of the hydrogen atom with fine structure one is interested in the eigenvalues of the Schrödinger radial equation with the potential[1,2]

$$V(r) = -(2/r + \omega/r^3).$$

Here $\omega = -\alpha^2\{j(j+1) - l(l+1) - \frac{3}{4}\}$, $\alpha = 1/137$ is Sommerfeld's fine structure constant and $j = l \pm \frac{1}{2}$. As ω is a small number, the eigenvalues of the problem

$$\frac{d^2\chi}{dr^2} + \left(2E + \frac{2}{r} - \frac{l(l+1)}{r^2} + \frac{\omega}{r^3}\right)\chi = 0, \qquad (1.1)$$

$$\int_0^\infty \chi^2(r)\,dr = 1$$

are usually (if $l > 0$) approximated by

$$E_{l,\nu} \approx -\frac{1}{2(l+\nu+1)^2} - \omega\int_0^\infty \frac{1}{r^3}[\chi_{l,\nu}(r)]^2\,dr \qquad (1.2)$$

where $\chi_{l,\nu}$ are the normalized eigenfunctions of the unperturbed potential ($\omega = 0$).

Now $\omega = \alpha^2(l+1)$ if $j = l - \frac{1}{2}$ and $\omega = -\alpha^2 l$ if $j = l + \frac{1}{2}$. In Sec. 9 we remark that for positive ω the spectrum consists of all E, and thus the formula (1.2) does not apply in this case.

In this paper we generalize the problem:

$$r^2\chi'' + \left(\left(\sum_{i=M}^{N} b_i r^i\right)\right)\chi = 0, \qquad (1.3)$$

$$M \leq 0 < N \quad \text{and} \quad b_M < 0 \text{ if } M < 0.$$

In the case of scattering problems we have $N = 2$ and $b_2 = 2E = k^2 > 0$. In the case of eigenvalue problems we have $N \geq 2$ and $b_2 = 2E$, E being the eigenvalue parameter, and $b_N < 0$. An example is the differential equation (1.1) with $N = 2$, $M = -1$, $b_1 = 2$, $b_0 = -l(l+1)$, and $b_{-1} = \omega$.

The differential equation (1.3) has two singular points: $r = \infty$ an irregular singular point with rank $N/2$, $r = 0$ is an irregular singular point with rank $|M|/2$, if $M < 0$ or $r = 0$ is a regular singular point, if $M = 0$.

Besides special cases, not discussed here, there are two power-series solutions of (1.3),

$$\chi_j(r) = r^{\rho_j}\sum_{n=-\infty}^{\infty} c_n^j r^n, \quad j = 1, 2, \qquad (1.4)$$

which are independent of each other and converge in the complex r plane, $0 < |r| < \infty$. If $r = 0$ is a regular singular point, then $c_n^j = 0$ for $n < 0$. Algorithms for computing the coefficients c_n^j and the characteristic exponents ρ_j were given in Ref. 3. and an algorithm for the special case that $\{b_i,\ i < 0\}$ are small numbers is given in Sec. 5 of this paper. It is useful for Eq. (1.1), because $b_{-1} = \omega$ is small.

Having found the series (1.4), most of the computing work has been done. It is easy to determine their asymptotic behavior at $r = \infty$ and $r = 0$, using the technique described in Ref. 4, which generalizes Kohno's work[5] to the case of two irregular singular points.

2. CONNECTION COEFFICIENTS

Specializing the rank = 1 at $r = \infty$, the method to compute the connection coefficients T_1 and T_2 in Eq. (2.1) becomes particularly easy. Because of its interest in scattering problems, where $N = 2$ and rank = 1, we will demonstrate the method.

A series solution $\chi(r) = \sum c_n r^{n+\rho}$ behaves asymptotically at $r \to \infty$ like

$$\chi(r) \sim T_1\chi_{\text{asy}}^1(r) + T_2\chi_{\text{asy}}^2(r), \qquad (2.1)$$

where

$$\chi_{\text{asy}}^k(r) = \exp(\lambda_k r)r^{\mu_k}\sum_{s=0}^{\infty} h_s^k r^{-s}, \quad k = 1, 2,$$

$$\lambda_1 = \sqrt{-b_2} = \sqrt{-2E}, \quad \text{Im}(\lambda_1) \geq 0,$$

$$\lambda_2 = -\lambda_1,$$

$$\mu_k = -b_1/(2\lambda_k).$$

The h_s are given by the recursion formula

$$2\lambda s h_s = h_{s-1}[(s-1-\mu)(s-\mu) + b_0] + \sum_{i=M}^{-1} b_i h_{s-1+i}$$

and $h_0 = 1$, $h_s = 0$ if $s < 0$. Here we have written λ, μ, h_s in place of λ_k, μ_k, h_s^k. Then we introduce

$$\delta_k = \rho - \mu_k, \quad g_n^k = (\lambda_k)^{n+\delta_k}/(n+\delta_k)!,$$

$$f_n^k = \sum_{s=0}^{\infty} g_{n+s}^k h_s^k, \quad k = 1, 2.$$

The larger n, the better this series will converge. The convergence rate is like that of a geometric series.

Finally we get the connection coefficients T_1 and T_2 by the two linear equations

 0022-2488/78/1906-1426$1.00

$$T_1 f_n^1 + T_2 f_n^2 = c_n,$$

$$T_1 f_{n+1}^1 + T_2 f_{n+1}^2 = c_{n+1}$$

for sufficiently large n.

We may take n so that (Ref. 4, p. 162)

$$\sum_{i=M}^{M} |b_i| < |(n+N+\rho)(n+N-1+\rho)|.$$

Example 2.1: Scattering problem for the Coulomb field: It is well known[6] that the solutions $\chi(r)$ of

$$\chi'' + \left(k^2 - \frac{2k\kappa}{r} - \frac{l(l+1)}{r^2}\right)\chi = 0,$$

which are regular at $r=0$, behave like

$$\chi(r) \sim \text{const} \times \sin[kr - \kappa \ln(2kr) - l\pi/2 + \eta_l], \quad r \to \infty, \tag{2.2}$$

where $\eta_l = ph[(l+i\kappa)!]$.

To apply our method, we start with the series solution

$$\chi(r) = r^\rho(1 + c_1 r + \cdots), \quad \rho = l+1,$$

$$\lambda_1 = ik = -\lambda_2, \quad \mu_1 = -i\kappa = -\mu_2,$$

$$\delta_1 = l+1+i\kappa, \quad \delta_2 = l+1-i\kappa,$$

and, dropping the index 1 or 2, we have $h_0 = 1$ and

$$h_s = h_{s-1}[s-1+\delta-(2\rho-1)](s-1+\delta)/(2\lambda s),$$

$$f_n = [\lambda^{n+\delta}/(n+\delta)!]F(-(2\rho-1)+\delta, \delta, n+1+\delta; \tfrac{1}{2}),$$

where F is the Gaussian hypergeometric series.

Using $F(a,b,b;x) = (1-x)^{-a}$, we get

$$f_{-1} = [(\lambda)^{-1+\delta}/(-1+\delta)!]2^{-2l-1+\delta}.$$

From the first equation of the linear system

$$T_1 f_{-1}^1 + T_2 f_{-1}^2 = c_{-1} = 0,$$

$$T_1 f_0^1 + T_2 f_0^2 = c_0 = 1,$$

we get just up to a constant A

$$T_1 = A \frac{(\lambda_2)^{-1+\delta_2}}{(-1+\delta_2)!} 2^{-1+\delta_2} \tag{2.3}$$

$$T_2 = -A \frac{(\lambda_1)^{-1+\delta_1}}{(-1+\delta_1)!} 2^{-1+\delta_1}.$$

An easy computation shows that

$$\chi(r) \sim T_1 \exp[ikr + \mu_1 \ln(r)] + T_2 \exp[-ikr + \mu_2 \ln(r)]$$

gives formula (2.2).

Example 2.2: Eigenvalues of the hydrogen atom: We consider solutions $\chi(r)$ of

$$\chi'' + \left(2E + \frac{2}{r} - \frac{l(l+1)}{r^2}\right)\chi = 0$$

that are regular at $r=0$. We have to find eigenvalues E for which a regular solution at $r=0$ also vanishes at infinity.

$$\chi(r) = r^{l+1}(1 + c_1 r + \cdots)$$

has the asymptotic behavior

$$\chi(r) \sim T_1 \exp(\lambda_1 r) r^{\mu_1}$$

$$\lambda_1 = \sqrt{-2E}, \quad \mu_1 = -1/\lambda_1 = -1/\sqrt{-2E} = -\mu_2.$$

λ_1 is positive, and we have to compute the zeros of $T_1(E)$.

According to Eq. (2.3), $T_1(E) = 0$ if $(-1+\delta_2)! = \infty$; hence, $\delta_2 = -\nu = 0, -1, -2, \cdots$, $l+1-\mu_2 = -\nu$ gives the well-known formula

$$E = -1/2(l+1+\nu)^2.$$

3. EIGENVALUE PROBLEM FOR THE CASE, WHEN $r=0$ IN AN IRREGULAR SINGULAR POINT

This case appears if $M < 0$ in Eq. (1.3), $b_M < 0$, and $b_N < 0$. At first we compute the series (1.4), $\chi_j(r)$, $j = 1, 2$. Having obtained the connection coefficients T_1^j, T_2^j we know $\chi_j(r) \sim T_1^j \cdot \chi_{\text{asy}}^1(r)$ for $r \to \infty$. Hence we have found the combination

$$A_1^+ \chi_1(r) + A_2^+ \chi_2(r) \sim 0, \quad r \to \infty, \tag{3.1}$$

where $A_1^+ = T_1^2$, $A_2^+ = -T_1^1$.

The transformation $r = 1/s$ (or $r = 1/s^2$, if the rank $|M|/2$ is not an integer), applied to the differential equation (1.3), and the computation of the connection coefficients at $s = \infty$ give us the right combination of $\chi_1(r)$ and $\chi_2(r)$ at $r=0$:

$$A_1^- \chi_1(r) + A_2^- \chi_2(r) \sim 0, \quad r \to 0. \tag{3.2}$$

E is eigenvalue, if both combinations (3.1) and (3.2) agree; hence the eigenvalue condition is

$$A_1^+/A_2^+ = A_1^-/A_2^-. \tag{3.3}$$

For example (1.1) these two ratios are plotted as functions of E in Fig. 1, and an analytic approximation is given in Sec. 8.

4. SCATTERING PROBLEM FOR THE CASE, WHEN $r=0$ IS AN IRREGULAR SINGULAR POINT

This case appears, if $M < 0$ in Eq. (1.3), $b_M < 0$ and $N = 2$, $b_2 = 2E$ is positive. For a given positive E we first compute the combination (3.2), which vanishes at $r=0$. The scattering phase shift then follows from the asymptotic behavior of this combination at $r = \infty$.

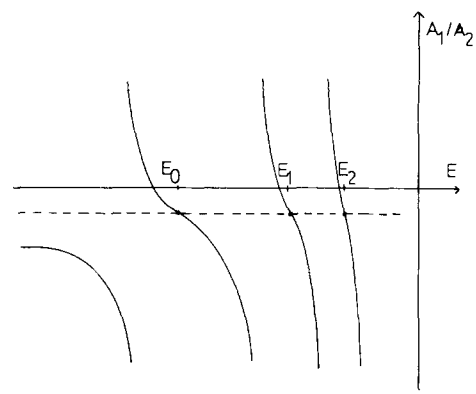


FIG. 1. Computation of eigenvalues for Example 8.1, by the method of Sec. 3: A_1^+/A_2^+ _____ dark line; A_1^-/A_2^- - - - - - dotted line. These ratios are computed as functions of E.

5. COMPUTATION OF POWER SERIES SOLUTIONS

Iteration methods for computing the series solutions (1.4) can be found in Ref. 3. If the coefficients b_i with $i<0$ of the differential equation (1.3) are small, these methods can be improved, as is done in this section.

As an example we treat the differential equation (1.1). Using $x(r)=r^{-1/2}\chi(r)$, we transform (1.1) into

$$x''+\frac{1}{r}x'+\left(2E+\frac{2}{r}-\frac{(l+1/2)^2}{r^2}+\frac{\omega}{r^3}\right)x=0. \tag{5.1}$$

More generally we transform (1.3) into

$$r^2x''+rx'+(\sum \tilde{b}_i r^i)x=0.$$

with

$$\tilde{b}_i=b_i \text{ if } i\neq 0 \text{ and } \tilde{b}_0=b_0-\tfrac{1}{4}.$$

For the purpose of the following sections we write b_i instead of $\tilde{b}_i$.

We treat the differential equation

$$r^2x''+rx'+\sum_{M}^{N} b_i r^i\ x=0 \tag{5.2}$$

and compute the series

$$x(r)=r^{\rho}\sum_{-\infty}^{\infty} c_n r^n. \tag{5.3}$$

A substitution $r\to ar$ changes $b_i\to b_i/a^i$ and $c_n\to c_n/a^n$, while b_0 and the characteristic exponent ρ remain unchanged. Then $\sum'_{i\neq 0}|b_i|$ is changed to

$$\epsilon=\sum'|b_i/a^i|$$

For the following method we choose a so that ϵ becomes small. This is possible if b_i are small numbers for negative i.

We now assume this transform has been done already and write

$$\epsilon=\sum'|b_i|.$$

For the following theory we use the following notation: $\delta_{i,j}$ Kronecker symbol,

$$\sigma=\pm 1,\quad \eta=\sigma\sqrt{-b_0},$$

$$R_d=\{\rho;\ |\rho-\eta|<d\},\quad \rho \text{ complex},$$

$$Q(\rho)=\max_{n\neq 0} 1/|(n+\rho)^2-\eta^2|,$$

$$\tilde{Q}(\rho)=\max_{n\neq 0}|n+\rho|/|[(n+\rho)^2-\eta^2]^2|.$$

With the differential equation (5.1) we have $\eta=\pm(l+\frac{1}{2})$ and $Q(\eta)=\infty$ if l is half-integer valued. But if 2η is no integer, then there is a positive d and

$$Q_d=\max_{\rho\in R_d}|Q(\rho)|<\infty,$$

$$\tilde{Q}_d=\max_{\rho\in R_d}|\tilde{Q}(\rho)|<\infty.$$

Inserting the series (5.3) into the differential equation (5.2) gives the following problem: Find (ρ,c); $c=\{c_n\}$, $c_0=1$,

$$\|c\|=\max_{-\infty<n<\infty}|c_n|<\infty$$

and

$$g_n(\rho,c):=[(n+\rho)^2+b_0]c_n+\sum_{i=M}^{N}{}' b_i c_{n-i}=0. \tag{5.4}$$

Lemma 1: There are constants a, b, so that the only bounded solutions (ρ,c) of (5.4) behave like

$$c_n=O(a^n/(n!)^{2/N})$$

$$c_{-n}=O(b^n/(n!)^{2/|M|})$$

for $n\to\infty$

Proof: This follows from a well-known theory of Perron–Kreuser difference equations (Ref. 7; see also Ref. 4, p. 171).

Corollary: Bounded solutions of (5.4) define convergent power series solutions (5.3).

Algorithm 1:

start: $c_n:=\delta_{n,0}$;

iterate:

$$c_n:=-(\sum' b_i c_{n-i})/[(n+\rho)^2+b_0] \tag{5.5}$$

$n:=1,-1,2,-2,\cdots$;

go to iterate;

end.

Lemma 2: If $2\eta\notin \mathbb{Z}\backslash\{0\}$, $\epsilon Q_d<1$, and $\rho\in R_d$, then Algorithm 1 converges and has a bounded solution $\{c_n(\rho)\}$,

$$\|c\|<1/(1-\epsilon Q_d).$$

Proof: Obvious. ||

In Algorithm 1 $c_0(\rho)=1$ and Eqs. (5.4) are considered for $n\neq 0$. We mean by $\tilde{g}_0(\rho)$ the value of $g_0(\rho,c(\rho))$. Hence Algorithm 1 defines a mapping

$$\rho\to\tilde{g}_0(\rho),$$

and we have to find the zeros of $\tilde{g}_0(\rho)$, because $\tilde{g}_0(\rho)=0$ if and only if ρ is the characteristic exponent.

We write $\tilde{g}_0(\rho)=\rho^2+b_0+\sum' b_i c_{-i}(\rho)=0$ in the form $\rho=T(\rho)$ with

$$T(\rho)=\sigma[-b_0-\sum' b_i c_{-i}(\rho)]^{1/2} \tag{5.6}$$

and look to see if T is a contraction mapping of R_d into R_d.

Theorem 1: If $2\eta\notin Z$ and ϵ is small enough to satisfy, with appropriate $d>0$, the conditions

$$\epsilon Q_d<1, \tag{5.7}$$

$$\epsilon/(1-\epsilon Q_d)<|b_0|, \tag{5.8}$$

$$\epsilon/(|\eta|(1-\epsilon Q_d))<d, \tag{5.9}$$

$$\epsilon^2\tilde{Q}_d/\left\{\min_{\rho\in R_d}|\rho|\cdot(1-\epsilon Q_d)^2\right\}<1, \tag{5.10}$$

then T is a constraction mapping of R_d into R_d. The iteration of T converges.

Proof: We first show that $T(\rho)\in R_d$ if $\rho\in R_d$. From $T(\rho)-\eta=\eta\{(1+[\sum' b_i c_{-i}(\rho)]/b_0)^{1/2}-1\}$ and the estimate

$$|(1+x)^{1/2}-1|\leq|x| \text{ if } |x|\leq 1,$$

we get, using (5.8),

$$|T(\rho)-\eta|<\epsilon/[|\eta|(1-\epsilon Q_d)];$$

hence, using (5.9), we have $T(\rho)\in R_d$.

Next we show that $|dT(\rho)/d\rho|<1$.

$$\frac{dT(\rho)}{d\rho}=\frac{-1}{2T(\rho)}\sum{}' b_i\frac{dc_{-i}(\rho)}{d\rho}.$$

Differentiating the equations $g_n(\rho,c(\rho))=0$, $n\neq 0$, we get an inhomogeneous system of linear equations for $dc_n/d\rho$:

$$\begin{aligned}&\frac{dc_0}{d\rho}=0,\\&\frac{dc_n}{d\rho}=\frac{-1}{(n+\rho)^2+b_0}\sum{}' b_i\frac{dc_{n-i}}{d\rho}-\gamma_n,\quad n\neq 0\end{aligned}\tag{5.11}$$

with

$$\gamma_n=2(n+\rho)c_n(\rho)/[(n+\rho)^2+b_0].$$

From (5.5) we get

$$\|\gamma\|\leq 2\epsilon\tilde{Q}_d/(1-\epsilon Q_d)$$

and conclude from the estimate

$$\left\|\frac{dc}{d\rho}\right\|\leq\frac{\|\gamma\|}{(1-\epsilon Q_d)}$$

that

$$\left|\frac{dT(\rho)}{d\rho}\right|\leq\frac{\epsilon^2\tilde{Q}_d}{|T(\rho)|(1-\epsilon Q_d)^2},$$

which is less than 1 by assumption (5.10). ||

Most of the problems occur if b_0 is negative, because Q_d then may become large or even infinite. η is real if b_0 is negative. We write

$$\eta=m/2+\Delta\eta,\quad m\in\mathbb{Z},$$
$$0\leq|\Delta\eta|\leq\tfrac{1}{4}.$$

Theorem 2: If 2η is not an integer, $|\eta|>\frac{3}{4}$, $\epsilon<\frac{1}{3}|\Delta\eta|$, and d is chosen to be $d=\frac{2}{3}|\Delta\eta|$, then T is a contraction mapping of R_d into R_d.

Proof: We introduce $\Delta\rho$ by $\rho=\eta+\Delta\rho$, and get

$$\begin{aligned}(n+\rho)^2-\eta^2&=((n+\Delta\rho)+\eta)^2-\eta^2\\&=(n+\Delta\rho)(n+\Delta\rho+2\eta)\\&=(n+\Delta\rho)(n+m+2\Delta\eta+\Delta\rho).\end{aligned}$$

If $\rho\in R_d$, we have $|\Delta\rho|\leq\frac{2}{3}|\Delta\eta|$. Using $|\eta|>\frac{3}{4}$, hence $|m|\geq 2$, we get

$$Q_d<\tfrac{2}{3}/|\Delta\eta|\quad\text{and}\quad\tilde{Q}_d<1/|\Delta\eta|^2.$$

With these estimates the conditions of Theorem 1 are fulfilled. ||

The method can be summarized by the following:

Algorithm 2 (to be used if 2η is not an integer):

start:

$\sigma:=\pm 1$, $\eta:=\sigma\sqrt{-b_0}$,

$\rho:=\eta$, $c_n:=\delta_{n,0}$;

inner loop:

$c_n:=(-\sum{}' b_ic_{n-i})/[(n+\rho)^2+b_0]$,

$n:=1,-1,2,-2,\cdots$;

outer loop: $\rho:=T(\rho)$;

end.

We take the sign $\sigma=1$ or $\sigma=-1$ of the square root in (5.6) during the whole iteration process and thus get two independent solutions, if the characteristic exponents do not differ by an integer. If $b_i=0$ for negative i, then the characteristic exponents are $\eta=\sigma\sqrt{-b_0}$ and Algorithm 2 without the outer loop is the usual way of computing the series solutions (5.3), setting $c_n=0$ for negative n.

We now consider the case that b_0 is close to zero. If $b_0\approx 0$, we do not iterate the operator $T(\rho)$, but we solve $\tilde{g}_0(\rho)=0$ by Newton's method: Replace

$$\rho\to\rho-\tilde{g}_0(\rho)\Big/\frac{d\tilde{g}_0(\rho)}{d\rho}.$$

We should no longer use η as a starting value for ρ; instead we notice that

$$\rho\approx\sigma(-b_0+\sum{}' b_ib_{-i}/i^2)^{1/2}$$

is a better approximation if $b_0\approx 0$, as will be shown in Eq. (6.1).

The method can be summarized by the following:

Algorithm 3 (to be used if $b_0\approx 0$):

start:

$\sigma:=\pm 1$, $\rho:=\sigma(-b_0+\sum{}' b_ib_{-i}/i^2)^{1/2}$,

$c_n:=\delta_{n,0}$, $c'_n:=0$;

inner loop:

$c_n:=(-\sum{}' b_ic_{n-i})/[(n+\rho)^2+b_0]$,

$c'_n:=-[\sum{}' b_ic'_{n-i}+2(n+\rho)c_n]/[(n+\rho)^2+b_0]$,

$n:=1,-1,2,-2,\cdots$;

outer loop:

$g_0:=\rho^2+b_0+\sum{}' b_ic_{-i}$,

$g'_0:=2\rho+\sum{}' b_ic'_{-i}$,

$\rho:=\rho-g_0/g'_0$;

end.

Now we treat the case, that 2η is close to an integer $m\neq 0$. If $2\eta\approx m\neq 0$, then the denominator in (5.5) will become small or even zero at $n=-m$ and the assumptions of Theorem 1 are hurt. We then modify the algorithm 2:

Algorithm 4: (to be used, if $2\eta\approx m\neq 0$): One has to compute the zeros of $G(c_{-m})$, which is the result of the following procedure.

procedure: $G(c_{-m})$;

start:

$\sigma:=\pm 1$, $\eta:=\sigma\sqrt{-b_0}$,

$\rho:=\eta$, $c_n:=\delta_{n,0}+c_{-m}\delta_{n,-m}$;

inner loop:

$c_n:=(-\sum{}' b_ic_{n-i})/((n+\rho)^2+b_0)$,

if $n\neq 0$ and $n\neq -m$.

outer loop: $\rho := T(\rho)$;

result: $G := g_{-m}(\rho, c)$;

end.

6. AN APPROXIMATION OF THE POWER SERIES SOLUTIONS

We assume that 2η is different from $\pm 1, \pm 2, \cdots$ and $\epsilon Q_d < 1$. We notice that

$$c_{-k} = -\sum{}' b_i c_{-k-i}/[(-k+\rho)^2 + b_0] = O(\epsilon) \quad (k \neq 0),$$

because the denominator does not become small.

Hence, using $c_0 = 1$, we get

$$c_{-k} \approx -b_{-k}/[(-k+\rho)^2 + b_0]$$

and from

$$\rho = \sigma(-b_0 - \sum{}' b_k c_{-k})^{1/2}$$

we derive the following improvement for a first estimate of ρ:

$$\rho \approx \sigma\{-b_0 + \sum{}' b_k b_{-k}/[(-k+\eta)^2 + b_0]\}^{1/2}, \tag{6.1}$$

where $\eta = \sigma\sqrt{-b_0}$.

With this approximation (6.1) we now replace Eqs. (5.4) by

$$[(n+\rho)^2 - \rho^2]c_n + \sum{}' b_i c_{n-i} = 0. \tag{6.2}$$

This system has characteristic exponent $\pm\rho$. We notice that the bounded solutions of the original system (5.4) depend continously on b_0. Hence the solutions of the system (6.2) approximate those of (5.4), because $b_0 + \rho^2 = O(\epsilon^2)$.

We conclude from Lemma 1 that we may neglect the right-hand term of

$$[(n+\rho)^2 - \rho^2]c_n + \sum_{i=1}^{N} b_i c_{n-i} = \sum_{i=M}^{-1} b_i c_{n-i}$$

for $n > 0$. Thus we approximately split up the system (6.2) into

$$c_n^+ = -\sum_{i>0} b_i c_{n-i}^+/[(n+\rho)^2 - \rho^2], \quad n > 0,$$

$$c_{-n}^+ = 0$$

and into

$$c_n^- = -\sum_{i<0} b_i c_{n-i}^-/[(n+\rho)^2 - \rho^2], \quad n < 0,$$

$$c_{-n}^- = 0$$

with $c_0^\pm = 1$. The recurrence relations for $c_n^\pm$ can easily be computed. In Table I we show how good the approximation

$$c_n \approx c_n^+ \text{ if } n > 0,$$

$$c_n \approx c_n^- \text{ if } n < 0 \tag{6.3}$$

agrees with the exact solution of the original system.

Expressed in terms of differential equations, we had replaced

$$r^2 x'' + rx' + \{b_0 + \sum{}' b_i r^i\}x = 0 \tag{6.4}$$

by

TABLE I. Example 7.1 is computed with Algorithm 2 and the approximation of Sec. 6.

	Algorithm 2		Approximation
	1. step	4. step (exact sol.)	of Sec. 6
$\rho =$	1.2	1.200098	1.200098
$n =$			
−20	0	$0.479276E-64$	$0.479246E-64$
−10	0	$0.125559E-25$	$0.125549E-25$
−2	0	$0.699805E-3$	$0.699559E-3$
−1	0	$-0.199906E-1$	$-0.199972E-1$
0	1	1	1
1	0	$-0.588167E-2$	$-0.588201E-2$
2	0	$0.172977E-4$	$0.172990E-4$
10	0	$0.136660E-28$	$0.136615E-28$
20	0	$0.101006E-62$	$0.101018E-62$

$$r^2 y'' + ry' + \left(-\rho^2 + \sum_{i>0} b_i r^i\right)y = 0 \quad \text{for } |r| < 1 \tag{6.5}$$

and by

$$r^2 z'' + rz' + \left(-\rho^2 + \sum_{i<0} b_i r^i\right)z = 0 \quad \text{for } |r| < 1. \tag{6.6}$$

(6.4) has two irregular points. (6.5) and (6.6) only have one irregular singular point.

There are solutions

$$y(r) = r^\rho \sum_{n=0}^{\infty} c_n^+ r^n \tag{6.7}$$

and

$$z(r) = r^{-\rho} \sum_{n=-\infty}^{0} c_n^- r^n \tag{6.8}$$

of Eq. (6.5) and (6.6).

As $c_0^\pm = 1$ we finally approximate

$$x(r) \approx y(r) + z(r) - r^\rho. \tag{6.9}$$

We always use formula (6.1) for ρ.

7. EXAMPLES

We treat the differential equation (5.1) for the parameters E, l, and ω arbitrarily chosen, but ω small. As ω is small, we substitute $r \to 100r$ and thus get small numbers b_{-1}, b_1, and b_2, while b_0 remains unchanged.

Example 7.1: We choose $l = 0.7$, $\alpha = 0.02$, $E = -0.173$. After the substitution $r \to 100r$ we obtain with $\omega = -\alpha^2 l$

$$b_{-1} = -0.028, \quad b_0 = -1.44, \quad b_1 = 0.02, \quad \text{and}$$

$$b_2 = -0.0000346.$$

We apply Algorithm 2. $\eta = \pm\sqrt{-b_0} = \pm 1.2$ is the starting value for ρ. With $\operatorname{sgn}\sigma = 1$, the characteristic exponent has been computed in Table I to be $\rho = 1.200098$.

Example 7.2: The example

$$b_2 = -0.000025, \quad b_1 = 0.02,$$

$$b_0 = -2.25, \quad b_{-1} = -0.000106559$$

must be solved by Algorithm 4, because $\eta = \pm(l + 1/2) = \pm 1.5$ is half-integer valued. If $\eta = 1.5$, corresponding to $\sigma = 1$, we use Algorithm 4 with $m = -3$. The results are

1. solution: $\rho = 1.500017$, $c_{-3} = -0.002887$,

2. solution: $\rho = 1.499983$, $c_{-3} = 70.53$,

On the other hand, with $\eta = -1.5$ corresponding to $\sigma = -1$, we get $m = 3$ and

2. solution: $\rho = -1.500017$, $c_3 = 0.1432$,

which, of course, agrees with the former second solution up to a constant factor, because the characteristic exponents differ by an integer number.

We recall that the sum of the characteristic exponents of the first and the second solution is equal to an integer [see Ref. 4, formula (1.5)].

Example 7.3: We transform the differential equation (5.1) into

$$x'' + \frac{1}{r}x' + \left(-\frac{1}{4} + \frac{\kappa}{r} - \frac{(l+\frac{1}{2})^2}{r^2} + \frac{B}{r^3}\right)x = 0, \tag{7.1}$$

$$\kappa = 1/\sqrt{-2E}, \quad B := \omega\cdot\sqrt{-8E} < 0.$$

(7.1) is split up as done in Eqs. (6.5) and (6.6) into

$$y'' + \frac{1}{r}y' + \left(-\frac{1}{4} + \frac{\kappa}{r} - \frac{\rho^2}{r^2}\right)y = 0,$$

which is related to Whittakers differential equation, and into

$$z'' + \frac{1}{r}z' + \left(-\frac{\rho^2}{r^2} + \frac{B}{r^3}\right)z = 0,$$

which is related to the modified Bessel equation.

The solutions are

$$y(r) = r^{-1/2}M_{\kappa,\rho}(r) \sim r^{\rho} \text{ if } r \text{ is small}$$

and

$$z(r) = \frac{(2\rho)!\,2^{2\rho}}{(-4B)^{\rho}} I_{2\rho}\left[\left(\frac{-4B}{r}\right)^{1/2}\right] \sim r^{-\rho} \text{ if } r \text{ is large.}$$

(For the notation see Ref. 8, pp. 375, 505.) Hence the approximation (6.9) is

$$x_1(r) \approx r^{-1/2}M_{\kappa,\rho}(r) + (-2\rho)!\,2^{-2\rho}(-4B)^{\rho}I_{-2\rho}(\sqrt{-4B/r}) - r^{\rho} \tag{7.2}$$

and

$$x_2(r) \approx r^{-1/2}M_{\kappa,-\rho}(r) + (2\rho)!\,2^{2\rho}(-4B)^{-\rho}I_{2\rho}(\sqrt{-4B/r}) - r^{-\rho}.$$

8. APPROXIMATION OF EIGENVALUES

We treat the last example 7.3, and want to choose κ so that a linear combination of $x_1(r)$ and $x_2(r)$ vanishes at $r = 0$ and at $r = \infty$. We notice that, for $r \to \infty$ and $s \to \infty$, $M_{\kappa,\rho}(r)$ and $I_{2\rho}(s)$ grow exponentially fast, while the combination

$$\frac{(2\rho)!}{(\rho - \frac{1}{2} - \kappa)!}M_{\kappa,-\rho}(r) - \frac{(-2\rho)!}{(-\rho - \frac{1}{2} - \kappa)!}M_{\kappa,\rho}(r) \sim 0$$

and the combination

$$I_{-2\rho}(s) - I_{2\rho}(s) \sim 0.$$

Hence, according to the formula (7.2), the combination

$$\frac{(-2\rho)!}{(-\rho - \frac{1}{2} - \kappa)!}x_1(r) - \frac{(2\rho)!}{(\rho - \frac{1}{2} - \kappa)!}x_2(r)$$

has approximately the right asymptotic behavior for $r \to \infty$, and the combination

$$\frac{(2\rho)!\,2^{2\rho}}{(-4B)^{\rho}}x_1(r) - (-2\rho)!\,2^{-2\rho}(-4B)^{\rho}x_2(r)$$

has the right asymptotic behavior for $r \to 0$. $E = -1/(2\kappa^2)$ is the eigenvalue, if both combinations agree:

$$1/(\rho - \tfrac{1}{2} - \kappa)! = F(\kappa) \tag{8.1}$$

with

$$F(\kappa) := \left(\frac{(-2\rho)!}{(2\rho)!}\right)^2\left(\frac{-2\omega}{\kappa}\right)^{2\rho}\frac{1}{(-\rho - \frac{1}{2} - \kappa)!}.$$

As ω is a small number and $\rho > 0$, we have to first order approximation $F(\kappa) \approx 0$; hence from formula (8.1) we obtain

$$\kappa \approx \tfrac{1}{2} + \rho + \nu, \quad \nu = 0, 1, 2, \cdots.$$

We recall that $\rho \approx l + \frac{1}{2}$, so $\kappa \approx \nu + l + 1$, the well-known result in the case $\omega = 0$.

To get a better approximation of Eq. (8.1), we introduce $\Delta\kappa$ by $\kappa = \frac{1}{2} + \rho + \nu + \Delta\kappa$; then (8.1) reads

$$\frac{1}{(-\nu - 1 - \Delta\kappa)!} = F(\kappa). \tag{8.2}$$

For (8.2) we can write

$$\frac{1}{(-1-\Delta\kappa)!} = \frac{(-1)^{\nu}F(\kappa)}{(1+\Delta\kappa)\cdots(\nu+\Delta\kappa)}$$

if $\nu \geq 1$. Using

$$\frac{1}{(-1+\epsilon)!} = \epsilon(1+\gamma\epsilon) + O(\epsilon^3)$$

for small ϵ ($\gamma = 0.577\cdots$ is the Eulers constant), we get

$$\Delta\kappa = \frac{(-1)^{\nu+1}F(\kappa)}{(1-\gamma\Delta\kappa)(1+\Delta\kappa)\cdots(\nu+\Delta\kappa)}, \tag{8.3}$$

which we solve by iteration.

Example 8.1: We choose $l = 0.7$, $j = 1.2$, $\alpha = 0.02$. Substitution of $r \to 100r$ just gives Example 7.1, if we also choose $E = -0.173$.

We approximate ρ with the formula (6.1) and observe from that formula that ρ is nearly independent of E. We got the result $\rho \approx 1.200098$. Now we compute E by (8.2), (8.3).

1. order approximation: $\kappa = \frac{1}{2} + \rho + \nu = 1.700098 + \nu$.

For $\nu = 0$ we get from (8.3) $\Delta\kappa = 0.315\,E - 8$. This gives the first eigenvalue $E_0 = -0.17299$. Setting $\nu = 1$ and $\nu = 2$, we get the higher eigenvalues $E_1 = -0.036521$, $E_2 = -0.0226337$.

9. CONCLUDING REMARKS

The most important methods of computing eigenvalues are the Ritz method and the perturbation formula. The first has the disadvantage that higher eigenvalues are rather inaccurate. The second has the disadvantage that one is not sure about the validity of the perturbation method.

An example is the differential equation (1.1) with positive ω. For small r the solutions behave asymp-

totically like $\chi(r) \sim Ar^{3/4} \sin(\sqrt{4\omega/r} + \delta)$. From this we conclude, that the spectrum consists of all E, $E < 0$ giving bound states and $E > 0$ scattering states.

The reason is that the Hamilton operator is not bounded from below, if $\omega > 0$. This is easily seen by using the Ritz variational principle with the trial functions $\chi(r) = r^{1+\epsilon} \exp(-r)$; $\epsilon > 0$.

The method given in this paper has the advantages:

(1) One computer program serves for both, eigenvalue and scattering problems.

(2) The accuracy can be increased by working with more digits.

[1]S. Flügge, *Quantentheorie I* (Springer-Verlag, Berlin, 1964), p. 270.
[2]S. Flügge, *Practical Quantum Mechanics II* (Springer-Verlag, Berlin, 1971), p. 17.
[3]F. Naundorf, "Ein Verfahren zur Berechnung der charakteristischen Exponenten von linearen Differentialgleichungen zweiter Ordnung mit zwei stark singulären Stellen," Z. Angew. Math. Mech. **57**, 47-49 (1977).
[4]F. Naundorf, SIAM J. Math. Anal. **7**, 157—75 (1976).
[5]M. Kohno, Hiroshima Math. J. **4**, 293—338 (1974).
[6]S. Flügge, *Practical Quantum Mechanics I* (Springer-Verlag, Berlin, 1971), p. 294.
[7]O. Perron, Math. Z., 16—24 (1959).
[8]Abramowitz and Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

Transmission problems for the Helmholtz equation

R. Kress

Lehrstühle für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany

G. F. Roach

Department of Mathematics, University of Strathclyde, Glasgow, Scotland
(Received 31 July 1977)

A straightforward treatment of these problems is given which appears to avoid many of the previously encountered difficulties. Admittedly some generality is lost by assuming that the various associated parameters k_i are not space dependent within their respective domains of definition, D_i. Nevertheless, by means of the approach offered here, such problems can be analyzed in just one function space; more general existence and uniqueness theorems can be obtained; there is no need to regularize the operators involved; and, above all, the solutions can be expressed in terms of certain boundary integral equations which, computationally, offer good prospects.

1. INTRODUCTION

Transmission problems associated with the Helmholtz equation have already been investigated by a few authors and in this connection we particularly mention the works of Kittappa and Kleinman,[1] Kupradze,[2] Latz,[3] and Werner.[4,5] Although these authors discuss such problems in quite general terms, for instance most consider the case when the Helmholtz equation has a distinct, space dependent associated parameter k_i in each of a number of regions $D_i \subset \mathbb{R}^3$, nevertheless the treatment of the problem and the results obtained are not always in a form which is entirely suitable for applications. For example, in Ref. 5 existence and uniqueness is only established for those values of the transmission coefficient which are related in a certain way to the ratio k_i/k_j, $i \neq j$ and, moreover, the solution is obtained in a form which is often difficult to handle numerically. On the other hand, the results obtained in Ref. 1 are only valid for small quantities and furthermore the "converse argument"[6] necessary when using a Green's representation of the solution seems to be missing.

In this paper we present what is felt to be a more straightforward treatment of these problems which appears to avoid the difficulties mentioned above. Admittedly we lose a certain amount of generality by demanding that the various k_i are not space dependent in the regions D_i. Nevertheless we find that with our approach we can analyze the problem in just one function space, the space of continuous functions; obtain more general existence and uniqueness theorems; avoid having to regularize the operators involved[6] and, above all, express the solution in terms of certain boundary integral equations which, computationally, offer a very hopeful prospect.

For clarity we have confined our attention here to problems associated with just one interface. Generalizations to problems involving not only many interfaces but also specific boundary conditions can be modeled on the work in Ref. 6.

2. STATEMENT OF THE PROBLEM

Let D_1 denote an unbounded, open region and D_2 a bounded open region of $\mathbb{R}^3$ such that

$$D_1 \cap D_2 = \emptyset, \quad \bar{D}_1 \cup D_2 = D_1 \cup \bar{D}_2 = \mathbb{R}^3.$$

The interface boundary $S := \partial D_1 = \partial D_2$ is assumed to be a closed, bounded Lyapunov surface. We shall denote by n the unit normal to S drawn in the direction from D_2 to D_1.

In this paper we shall be concerned with the following problem.

Transmission Problem: Find functions $u_m \in C^2(D_m) \cap C^1(\bar{D}_m)$, $m = 1, 2$ such that

$$\Delta u_m + k_m^2 u_m = 0, \quad \text{in } D_m, \quad m = 1, 2, \tag{2.1}$$

$$\left.\begin{aligned} \mu_1 u_1 - \mu_2 u_2 &= f \\ \frac{\partial u_1}{\partial n} - \frac{\partial u_2}{\partial n} &= g \end{aligned}\right\} \text{in } S, \tag{2.2}$$

where u_1 is required to satisfy a radiation condition of the form

$$u = O(1/R), \quad \partial u/\partial n - iku = o(1/R), \quad R \to \infty, \tag{2.3}$$

and $f \in C^{1,\alpha}(S)$, $g \in C^{0,\alpha}(S)$, and k_1, k_2, μ_1, μ_2 are nonzero complex numbers with $0 \leq \arg k_1$, $\arg k_2 < \pi$.

We note that the transmission conditions, (2.2), are quite typical in the sense that they can always be easily renormalized to read

$$u_1 - u_2 = \tilde{f}, \quad \lambda_1 \frac{\partial u_1}{\partial n} - \lambda_2 \frac{\partial u_2}{\partial n} = \tilde{g}$$

for appropriate functions $\tilde{f}$, $\tilde{g}$ and complex numbers λ_1, λ_2.

3. UNIQUENESS

For our immediate convenience we establish a uniqueness theorem for the homogeneous transmission problem

$$\Delta v_m^2 + k_m^2 v_m = 0, \quad \text{in } D_m, \quad m = 1, 2, \tag{3.1}$$

$$\left.\begin{aligned} \mu_1 v_1 - \mu_2 v_2 &= 0 \\ \frac{\partial v_1}{\partial n} - \frac{\partial v_2}{\partial n} &= 0 \end{aligned}\right\} \text{in } S, \tag{3.2}$$

where v_1 is required to satisfy a radiation condition of the form

 0022-2488/78/1906-1433$1.00

$$v = O\left(\frac{1}{R}\right), \quad \frac{\partial v}{\partial n} - ikv = o\left(\frac{1}{R}\right), \quad R \to \infty. \tag{3.3}$$

Theorem 3.1: Let $k_1, k_2 \in \mathbb{C}\backslash\{0\}$ with $0 \leq \arg k_1$, $\arg k_2 < \pi$, and let $\mu_1, \mu_2 \in \mathbb{C}\backslash\{0\}$ be such that

$$\rho := \frac{\mu_2 \bar{k}_2^2}{\mu_1 \bar{k}_1^2} \in \mathbb{R},$$

where $\rho \geq 0$ (<0) if $\mathrm{Re} k_1 \mathrm{Re} k_2 \geq 0$ (<0). Then the only solution of the homogeneous transmission problem is the trivial solution $v_1 = v_2 = 0$.

Proof: Let K_R be a sphere of radius R with outward drawn normal n. Assume that K_R contains S in its interior and let $D_{1,R} := \{x \in D_1 : |x| < R\}$. Applying Green's theorems over $D_{1,R}$ and D_2 and using (3.1) and (3.2) we obtain

$$\begin{aligned}\mu_1 \int_{K_R} v_1 \frac{\partial \bar{v}_1}{\partial n}\, ds &= \mu_1 \int_{D_{1,R}} |\operatorname{grad} v_1|^2 dx + \mu_2 \int_{D_2} |\operatorname{grad} v_2|^2 dx \\ &\quad - \mu_1 \bar{k}_1^2 \int_{D_{1,R}} |v_1|^2 dx - \mu_2 \bar{k}_2^2 \int_{D_2} |v_2|^2 dx. \end{aligned} \tag{3.4}$$

Dividing throughout (3.4) by $\mu_1 \bar{k}_1^2$ and taking imaginary parts we get

$$\begin{aligned}\mathrm{Im}\left(\frac{1}{\bar{k}_1^2} \int_{K_R} v_1 \frac{\partial \bar{v}_1}{\partial n}\, ds\right) &= \mathrm{Im}\left(\frac{1}{\bar{k}_1^2} \int_{D_{1,R}} |\operatorname{grad} v_1|^2 dx\right) \\ &\quad + \rho\, \mathrm{Im}\left(\frac{1}{\bar{k}_2^2} \int_{D_2} |\operatorname{grad} v_2|^2 dx\right). \end{aligned} \tag{3.5}$$

Since the integrals on the right-hand side are clearly real we need only concern ourselves with the complex nature of k_1 and k_2. Specifically we distinguish between two cases; $\mathrm{Im} k_1 > 0$ and $\mathrm{Im} k_1 = 0$.

In the case $\mathrm{Im} k_1 > 0$ and since v_1 satisfies the radiation condition it follows[2] that the left-hand side of (3.5) tends to zero as $R \to \infty$. Consequently (3.5) reduces to

$$\frac{\mathrm{Re} k_1 \mathrm{Im} k_2}{|k_1|^4} \int_{D_1} |\operatorname{grad} v_1|^2 dx + \rho \frac{\mathrm{Re} k_2 \mathrm{Im} k_2}{|k_2|^4} \int_{D_2} |\operatorname{grad} v_2|^2 dx = 0.$$

If $\mathrm{Re} k_1 \neq 0$, then from the sign of ρ it follows that $\int_{D_1} |\operatorname{grad} v_1|^2 dx = 0$. From this fact we conclude that $v_1 = 0$ in D_1. The transmission conditions now enable us to write $v_2 = \partial v_2 / \partial n = 0$ on S and as a consequence of Green's representation theorem it follows that $v_2 = 0$ in D_2.

If $\mathrm{Re} k_1 = 0$, $\mathrm{Re} k_2 \neq 0$, then it follows that $\int_{D_2} |\operatorname{grad} v_2|^2 dx = 0$; from which we conclude that $\operatorname{grad} v_2 = 0$ in D_2. The transmission conditions then show that $\partial v_1 / \partial n = 0$ on S and consequently, by the uniqueness theorem for the exterior Neumann problem we deduce that $v_1 = 0$ in D_1. Using the transmission conditions once again it follows that $v_2 = 0$ in D_2.

If $\mathrm{Re} k_1 = \mathrm{Re} k_2 = 0$, then $\bar{k}_1^2$ and $\bar{k}_2^2$ are negative and therefore from (3.4), since $\rho \geq 0$ we obtain

$$\int_{D_1} |v_1|^2 dx = \int_{D_2} |v_2|^2 dx = 0$$

which implies that $v_1 = v_2 = 0$.

In the case when $\mathrm{Im} k_1 = 0$ we obtain from (3.5)

$$\mathrm{Im} \int_{K_R} v_1 \frac{\partial \bar{v}_1}{\partial n}\, ds \geq 0.$$

From the radiation condition it follows that, as $R \to \infty$,

$$k_1 \int_{K_R} |v_1|^2 ds + \mathrm{Im} \int_{K_R} v_1 \frac{\partial \bar{v}_1}{\partial n}\, ds = o(1).$$

Since both terms on the left-hand side are nonnegative it follows that

$$\int_{K_R} |v_1|^2 ds = o(1)$$

as $R \to \infty$. Therefore, by Rellich's Theorem it follows that $v_1 = 0$ in D_1. The proof of the theorem is now completed by means of arguments similar to those used above.

We note that this theorem is a special case of a theorem obtained by Werner.[4]

It is obvious that by a more detailed discussion of (3.4), uniqueness can be established for a wider range of values of k_1, k_2, μ_1, and μ_2 than those considered in Theorem 3.1. However, the present range is sufficient for our purposes as it indicates a region in which uniqueness can be assured. That uniqueness cannot be expected in general for all values of k_1, k_2, μ_1, and μ_2 is illustrated by the following example.

Let D_1 be the exterior and D_2 the interior of the unit sphere in $\mathbb{R}^3$. Introduce spherical polar coordinates $(r, \theta, \phi) \equiv (r, \Theta)$ and consider the functions

$$u_{1,n,i}(r, \Theta) := a_{1,n,i} h_n^{(1)}(k_1 r) Y_{n,i}(\Theta), \quad 1 \leq r < \infty,$$

$$u_{2,n,i}(r, \Theta) := a_{2,n,i} j_n(k_2 r) Y_{n,i}(\Theta), \quad 0 \leq r \leq 1,$$

where $h_n^{(1)}$ and j_n denote the nth spherical Hankel function of the first kind and spherical Bessel function respectively and where $Y_{n,i}$, $i = 1, 2, \ldots, i_n$ denote the linearly independent spherical harmonics of order n in $\mathbb{R}^3$. Obviously these functions form a solution of the homogeneous transmission problem if and only if

$$\mu_1 a_{1,n,i} h_n^{(1)}(k_1) - \mu_2 a_{2,n,i} j_n(k_2) = 0,$$

$$k_1 a_{1,n,i} h_n^{(1)\prime}(k_1) - k_2 a_{2,n,i} j_n'(k_2) = 0.$$

This system of equations has a nontrivial solution for $a_{1,n,i}$ and $a_{2,n,i}$ if and only if the determinant

$$\begin{vmatrix} \mu_1 h_n^{(1)}(k_1) & \mu_2 j_n(k_2) \\ k_1 h_n^{(1)\prime}(k_1) & k_2 j_n'(k_2) \end{vmatrix} = 0.$$

That is, there are nontrivial solutions if and only if, given k_1, k_2 there holds

$$\frac{\mu_1}{\mu_2} = \frac{k_1 h_n^{(1)\prime}(k_1) j_n(k_2)}{k_2 h_n^{(1)}(k_1) j_n'(k_2)}.$$

Obviously for each k_1, k_2 there exists, in general, an infinite number of (complex) transmission coefficients (μ_1/μ_2) such that we have nonuniqueness.

4. EXISTENCE

We shall settle the question of existence of solutions of the transmission problem by establishing that Fredholm's Alternative is available for the problem.

Let γ_m, $m=1,2$, denote the fundamental solutions

$$\gamma_m(x,x') := -\frac{1}{2\pi}\frac{\exp(ik_m|x-x'|)}{|x-x'|}, \quad x\neq x'$$

of the appropriate Helmholtz equation. We define potentials

$$u_m(x) := \int_S \left(\phi(x')\frac{\partial\gamma_m}{\partial n(x')} + \psi(x')c_m\gamma_m\right)ds(x'),$$
$$x\in \mathbb{R}^3\backslash S, \quad m=1,2, \tag{4.1}$$

with densities $\phi\in C^{1,\alpha}(S)$, $\psi\in C^{0,\alpha}(S)$, and where $c_1, c_2 \in \mathbb{C}$ are fixed constants which will be appropriately chosen later.

From the regularity properties of single and double layer potentials[7,8] it follows that the potentials defined in (4.1) are such that $u_m\in C^2(D_m)\cap C^1(\bar{D}_m)$, $m=1,2$. Furthermore the potentials u_m are known to satisfy the differential equation (2.1) and the radiation condition (2.3).

Using the jump conditions of potential theory[7] we obtain on S

$$(\mu_1u_1-\mu_2u_2)(x) = \int_S\left[\phi(x')\left(\mu_1\frac{\partial\gamma_1}{\partial n(x')}-\mu_2\frac{\partial\gamma_2}{\partial n(x')}\right) + \psi(x')(\mu_1c_1\gamma_1-\mu_2c_2\gamma_2)\right]ds(x') - (\mu_1+\mu_2)\phi(x), \tag{4.2}$$

$$\left(\frac{\partial u_1}{\partial n}-\frac{\partial u_2}{\partial n}\right)(x) = \int_S\left[\phi(x')\left(\frac{\partial^2\gamma_1}{\partial n(x)\,\partial n(x')}-\frac{\partial^2\gamma_2}{\partial n(x)\,\partial n(x')}\right) + \psi(x')\left(c_1\frac{\partial\gamma_1}{\partial n(x)}-c_2\frac{\partial\gamma_2}{\partial n(x)}\right)\right]ds(x') + (c_1+c_2)\psi(x). \tag{4.3}$$

In both (4.2) and (4.3) we emphasize that $x\in S$.

We now introduce integral operators $A_{mj}: C(S)\to C(S)$, $m,j=1,2$ defined by

$$(A_{11}\phi)(x) := \int_S \phi(x')\left(\mu_1\frac{\partial\gamma_1}{\partial n(x')}-\mu_2\frac{\partial\gamma_2}{\partial n(x')}\right)ds(x'),$$
$$(A_{12}\psi)(x) := \int_S \psi(x')[\mu_1c_1\gamma_1-\mu_2c_2\gamma_2]\,ds(x'),$$
$$(A_{21}\phi)(x) := \int_S \phi(x')\left(\frac{\partial^2\gamma_1}{\partial n(x)\,\partial n(x')}-\frac{\partial^2\gamma_2}{\partial n(x)\,\partial n(x')}\right)ds(x'),$$
$$(A_{22}\psi)(x) := \int_S \psi(x')\left(c_1\frac{\partial\gamma_1}{\partial n(x)}-c_2\frac{\partial\gamma_2}{\partial n(x)}\right)ds(x').$$

In terms of these operators we have the obvious result:

Theorem 4.1: The potentials u_1, u_2 defined in (4.1) solve the transmission problem (2.1)–(2.3) if the densities ϕ and ψ solve the system of boundary integral equations

$$(\mu_1+\mu_2)\phi - A_{11}\phi - A_{12}\psi = -f,$$
$$(c_1+c_2)\psi + A_{21}\phi + A_{22}\psi = g. \tag{4.4}$$

The integral operators in (4.4) have weakly singular kernels. For the operators A_{11}, A_{12}, A_{22} this fact follows by standard potential theoretical arguments. For the operator A_{21} it is established by writing γ_m in the expanded form

$$\gamma_m(x,x') = \frac{1}{|x-x'|} + \sum_{j=1}^{\infty}\frac{(ik_m)^j|x-x'|^{j-1}}{j!}$$

and noticing that

$$\frac{\partial^2\gamma_1}{\partial n(x)\,\partial n(x')}-\frac{\partial^2\gamma_2}{\partial n(x)\,\partial n(x')} = O\left(\frac{1}{|x-x'|}\right).$$

Therefore, with $C(S)$ endowed with the usual supremum norm, it follows that the operators $A_{mj}: C(S)\to C(S)$ are compact, and consequently the Riesz–Schauder theory is available for our use.

We now introduce operators $E, A: C(S)\times C(S)\to C(S)\times C(S)$ defined by

$$E := \begin{bmatrix}(\mu_1+\mu_2)I & 0\\ 0 & (c_1+c_2)I\end{bmatrix}, \quad A := \begin{bmatrix}A_{11} & A_{12}\\ -A_{21} & -A_{22}\end{bmatrix},$$

where I is the identity operator on $C(S)$. Setting $\Phi := \begin{bmatrix}\phi\\ \psi\end{bmatrix}$, the integral equations (4.4) can be written in the form

$$(E-A)\Phi = F, \tag{4.5}$$

where $F := \begin{bmatrix}-f\\ g\end{bmatrix}$.

Remark: We shall always require the operator E to be invertible. Therefore, we must always have $\mu_1+\mu_2\neq 0$ and $c_1+c_2\neq 0$. The latter condition is a technical one which we can always satisfy. The former condition however is of a more intrinsic nature in that it influences the acceptable transmission conditions.

Theorem 4.2: If the homogeneous transmission problem (3.1)–(3.3) has only the trivial solution, then the inhomogeneous transmission problem (2.1)–(2.3) has a unique solution.

Proof: (i) Assume that the homogeneous integral equation

$$(E-A)\Phi = 0 \tag{4.6}$$

admits only the trivial solution. [That this is indeed the case will be proved in (ii)]. Then by the Riesz–Schauder theory the inhomogeneous equation (4.5) has a unique solution.

From the weak singularity of all the kernels involved it follows by standard arguments that

$$A_{mj}: C(S)\to C^{0,\alpha}(S), \quad m,j=1,2$$

and

$$A_{mj}: C^{0,\alpha}(S)\to C^{1,\alpha}(S), \quad m,j=1,2.$$

Therefore, since $f\in C^{1,\alpha}(S)$, $g\in C^{0,\alpha}(S)$ it automatically follows that for any continuous solution of (4.5) we have $\phi\in C^{1,\alpha}(S)$ and $\psi\in C^{0,\alpha}(S)$. Therefore, by means of Theorem 4.1, we see that from the solution of the inhomogeneous integral equation (4.5) we obtain a solution of the inhomogeneous transmission problem (2.1)–(2.3).

(ii) Let $\Phi := \begin{bmatrix}\phi\\ \psi\end{bmatrix}$ be a solution of the homogeneous integral equation (4.6). Then the potentials u_1, u_2, each with densities ϕ and ψ as in (4.1), solve the homogeneous transmission problem. Now by assumption, $u_m=0$ in D_m, $m=1,2$ whilst the jump conditions yield

$$u_m^+ - u_m^- = -2\phi, \quad \text{on } S, \quad m=1,2,$$

$$\frac{\partial u_m^+}{\partial n} - \frac{\partial u_m^-}{\partial n} = 2c_m\psi, \quad \text{on } S, \quad m=1,2.$$

Here by the indices + and − we distinguish between the values obtained by approaching S from inside D_1 and from inside D_2 respectively. Consequently we have

$$u_2^+ + u_1^- = 0 \quad \text{on } S,$$

$$\frac{1}{c_2}\frac{\partial u_2^+}{\partial n} + \frac{1}{c_1}\frac{\partial u_1^-}{\partial n} = 0 \quad \text{on } S.$$

If now we define

$$v_2 := \frac{1}{c_2}u_2 \ \text{ in } D_1, \quad v_1 := -\frac{1}{c_1}u_1 \ \text{ in } D_2,$$

then we see that v_1, v_2 satisfies the homogeneous transmission problem

$$\left.\begin{aligned} &\Delta v_m + k_m^2 v_m = 0 \ \text{ in } D_{m-1}, \quad m=1,2,\\ &c_2 v_2 - c_1 v_1 = 0\\ &\frac{\partial v_2}{\partial n} - \frac{\partial v_1}{\partial n} = 0 \end{aligned}\right\} \text{in } S,$$

v_2 satisfies the apporpriate radiation condition at infinity.

Notice that here we have set $D_0 \equiv D_2$.

Since the constants c_1, c_2 are at our disposal we choose them to ensure that

$$\eta := \frac{c_1\bar{k}_1^2}{c_2\bar{k}_2^2} \in \mathbf{R},$$

where $\eta \geq 0$ (<0) if $\mathrm{Re}k_1\,\mathrm{Re}k_2 \geq 0$ (<0). Then by the uniqueness theorem, Theorem 3.1, it follows that $v_2 = 0$ in D_1 and $v_1 = 0$ in D_2. Hence, by using the jump condition, we conclude that $\phi = \psi = 0$, and the proof of the theorem is complete.

In order to deal with the case when the homogeneous transmission problem has nontrivial solutions we must introduce an operator B which, with respect to some suitable structure, is an adjoint of the operator A. To this end define the bilinear form

$$(\cdot,\cdot) : (C(S)\times C(S))\times(C(S)\times C(S)) \to \mathbb{C} \text{ by}$$

$$(\Phi,\Psi) := \int_S (\phi\chi + \psi\eta)\,ds,$$

where $\Phi := [\phi\ \psi]$ and $\Psi := [\chi\ \eta]$. With respect to this bilinear form the operator E is seen to be self-adjoint whilst the adjoint of A is given by

$$B := \begin{bmatrix} B_{11} & -B_{12} \\ B_{21} & -B_{22} \end{bmatrix},$$

where

$$(B_{11}\chi)(x) := \int_S \chi(x')\left(\mu_1\frac{\partial\gamma_1}{\partial n(x)} - \mu_2\frac{\partial\gamma_2}{\partial n(x)}\right)ds(x'),$$

$$(B_{12}\eta)(x) := \int_S \eta(x')\left(\frac{\partial^2\gamma_1}{\partial n(x)\,\partial n(x')} - \frac{\partial^2\gamma_2}{\partial n(x)\,\partial n(x')}\right)ds(x'),$$

$$(B_{21}\chi)(x) := \int_S \chi(x')(\mu_1 c_1\gamma_1 - \mu_2 c_2\gamma_2)\,ds(x'),$$

$$(B_{22}\eta)(x) := \int_S \eta(x')\left(c_1\frac{\partial\gamma_1}{\partial n(x')} - c_2\frac{\partial\gamma_2}{\partial n(x')}\right)ds(x').$$

Let V be the linear space of all solutions of the homogeneous transmission problem (3.1)–(3.3) and let

$$W := \left\{ \Psi = \left[\begin{matrix} \frac{\partial v_1}{\partial n} \\ \mu_1 v_1 \end{matrix}\right]\Bigg|_S = \left[\begin{matrix} \frac{\partial v_2}{\partial n} \\ \mu_2 v_2 \end{matrix}\right]\Bigg|_S : (v_1, v_2) \in V \right\}.$$

Theorem 4.4: If $N(E-B)$ denotes the null space of the operator $E-B$, then $N(E-B) = W$.

Proof: (i) Let $(v_1, v_2) \in V$. By means of Green's representation we obtain

$$\int_S \left(\frac{\partial v_1(x')}{\partial n(x')}\gamma_1 - v_1(x')\frac{\partial\gamma_1}{\partial n(x')}\right)ds(x') = 2v_1(x), \quad x \in D_1, \tag{4.7}$$

$$\int_S \left(\frac{\partial v_2(x')}{\partial n(x')}\gamma_2 - v_2(x')\frac{\partial\gamma_2}{\partial n(x')}\right)ds(x') = 0, \quad x \in D_1. \tag{4.8}$$

Multiply (4.7) by $c_1\mu_1$ and (4.8) by $c_2\mu_2$, subtract the resulting equations, and use the transmission conditions (3.2) to obtain

$$\int_S \left[\frac{\partial v_1(x')}{\partial n(x')}\{\mu_1 c_1\gamma_1 - \mu_2 c_2\gamma_2\} - \mu_1 v_1(x')\left(c_1\frac{\partial\gamma_1}{\partial n(x')} - c_2\frac{\partial\gamma_2}{\partial n(x')}\right)\right]ds(x') = 2\mu_1 c_1 v_1(x), \quad x \in D_1. \tag{4.9}$$

Letting $x \to S$ we obtain

$$B_{21}\left(\frac{\partial v_1}{\partial n}\Big|_S\right) - B_{22}(\mu_1 v_1|_S) = (c_1+c_2)\mu_1 v_1|_S. \tag{4.10}$$

Setting $c_1 = c_2 = 1$ in (4.9) and taking the normal derivatives of the resulting equation we obtain, on letting $x \to S$,

$$B_{11}\left(\frac{\partial v_1}{\partial n}\Big|_S\right) - B_{12}(\mu_1 v_1|_S) = (\mu_1+\mu_2)\frac{\partial v_1}{\partial n}\Big|_S \tag{4.11}$$

Obviously (4.10) and (4.11) imply that

$$(E-B)\Psi = 0,$$

where

$$\Psi := \left[\begin{matrix} \frac{\partial v_1}{\partial n} \\ \mu_1 v_1 \end{matrix}\right]\Bigg|_S.$$

Therefore

$$W \subset N(E-B).$$

(ii) Now let $\Phi = [\phi\ \psi]$ be a solution of the homogeneous equation $(E-A)\Phi = 0$. Then, as shown in the proof of Theorem 4.2, we have $\phi \in C^{1,\alpha}(S)$ and $\psi \in C^{0,\alpha}(S)$, and the corresponding potentials u_1, u_2 defined by (4.1) solve the homogeneous transmission problem.

Now define a mapping

$$T : N(E-A) \to N(E-B)$$

by

$$T : \Phi = \begin{bmatrix} \phi \\ \psi \end{bmatrix} \mapsto \left[\begin{matrix} \frac{\partial u_1}{\partial n} \\ \mu_1 u_1 \end{matrix}\right]\Bigg|_S.$$

Clearly $T\Phi \in W$. Further, T is a linear operator which moreover is injective since from $T\Phi = 0$ it follows that $\partial u_1/\partial n = u_1 = 0$ on S and therefore $u_1 = 0$ in D_1. From the transmission conditions we obtain $u_2 = \partial u_2/\partial n = 0$ on S. If we now repeat the arguments used in the second part of Theorem 4.2, then we find that $\phi = \psi = 0$ and we conclude that $\Phi = 0$. Finally, since A and B are compact,

Fredholm's Alternative enables us to assert that $\dim N(E-A)=\dim N(E-B)$ and according we can deduce that T is bijective.

Now let $\Psi := \begin{bmatrix} x \\ \eta \end{bmatrix} \in N(E-B)$ and define $\Phi := T^{-1}\Psi$. Then $\Psi = T\Phi \in W$ and hence $N(E-B) \subset W$.

The theorem is established by combining the results of (i) and (ii). An immediate corollary to this theorem is

Corollary 4.4: The space V of all solutions to the homogeneous transmission problem is finite dimensional.

In keeping with the results of classical Fredholm theory we establish the following.

Theorem 4.5: In order that the nonhomogeneous transmission problem (3.1)–(3.3) have a solution it is necessary and sufficient that

$$\int_S \left(f \frac{\partial v_1}{\partial n} - g\mu_1 v_1 \right) ds = 0$$

for all solutions v_1, v_2 of the corresponding homogeneous problem.

Proof (Necessity): Let u_1, u_2 be solutions of the nonhomogeneous problem we obtain by Green's theorem,

$$\int_S \left(f \frac{\partial v_1}{\partial n} - g\mu_1 v_1 \right) ds$$

$$= \int_S \left[(\mu_1 u_1 - \mu_2 u_2) \frac{\partial v_1}{\partial n} - \left(\frac{\partial u_1}{\partial n} - \frac{\partial u_2}{\partial n} \right) \mu_1 v_1 \right] ds,$$

$$\int_S \left(f \frac{\partial v_1}{\partial n} - g\mu_1 v_1 \right) ds$$

$$= \mu_1 \int_S \left(u_1 \frac{\partial v_1}{\partial n} - \frac{\partial u_1}{\partial n} v_1 \right) ds$$

$$- \mu_2 \int_S \left(u_2 \frac{\partial v_2}{\partial n} - \frac{\partial u_2}{\partial n} v_2 \right) ds$$

$$= \mu_1 \int_{K_R} \left(u_1 \frac{\partial v_1}{\partial n} - \frac{\partial u_1}{\partial n} v_1 \right) ds.$$

This last integral tends to zero as $R \to \infty$ by virtue of the radiation condition.

(Sufficiency): Let $\Psi \in N(E-B)$. Then by Theorem 4.3 we have

$$\Psi = \left[\begin{matrix} \frac{\partial v_1}{\partial n} \\ \mu_1 v_1 \end{matrix} \right]_S$$

for some element $(v_1, v_2) \in V$. Then

$$\int_S \left(f \frac{\partial v_1}{\partial n} - g\mu_1 v_1 \right) ds = 0$$

implies that, for Eq. (4.5) we have $(F, \Psi) = 0$. But by the Fredholm theory this ensures that (4.5) has a solution, a fact which ensures that the given transmission problem has a solution.

We can now summarize these several results in the following manner.

Theorem 4.6: *Either* the homogeneous transmission problem (3.1)–(3.3) has only the trivial solution in which case the nonhomogeneous transmission problem (3.1)–(3.3) has a unique solution for any inhomogeneous term,

or the homogeneous transmission problem has a finite number of linearly independent solutions and the nonhomogeneous transmission problem is solvable if and only if the inhomogeneous term satisfies the conditions of Theorem 4.5.

[1] R. Kittappa and R.E. Kleinman, "Acoustic scattering by penetrable homogeneous objects," J. Math. Phys. **16**, 421–432 (1975).

[2] W.D. Kupradse, *Randwertaufgaben der Schwingungstheorie und, Integraleichungen* (Deutscher Verlag der Wissenschaften, Berlin, 1956).

[3] N. Latz, "Electromagnetic diffraction by imperfectly dielectric wedges," J. Math. Anal. Appl. **43**, 373–387 (1973).

[4] P. Werner, "Zur mathematischen Theorie akustischen Wellenfelder," Arch. Rat. Mech. Anal. **6**, 231–260 (1960).

[5] P. Werner, "Bengungsprobleme der mathematischen Akustik," Arch. Rat. Mech. Anal. **12**, 115–184 (1963).

[6] R. Kress and G. F. Roach, "On mixed boundary value problems for the Helmholtz equation," Proc. R. Soc. Edin. A 77, 65–77 (1977).

[7] N.M. Günter, *Die Potentialtheorie und ihre Anwendungen auf Grundaufgaben der mathematischen Physik* (Teubner, Leipzig, 1957).

[8] C. Miranda, *Partial Differential Equations Equations of Elliptic Type* (Springer, Berlin, 1970).

Functional integrals over anticommuting variables and the one-dimensional scattering problem[a)]

Stuart Samuel

Department of Physics, University of California, Berkeley, California 94720
(Received 8 November 1977)

Functional integrals over anticommuting variables are used to obtain a scattering formula for smooth localized potentials in one dimension. Via a calculational trick, the functional integral is evaluated to obtain the transition matrix coefficients as an expansion in
$[k'(x)/k(x)]\,(k(x) = \{(2m/\hbar^2)[E - V(x)]\}^{1/2})$.
This expansion is shown to have a simple physical interpretation.

I. INTRODUCTION

The purpose of this paper is 2-fold. First, a new scattering formula is obtained for one-dimensional quantum mechanical problems. This formula [Eq. (II. 12)] is expressed in terms of a functional integral over anticommuting variables. In such a form it is not very useful since functional integrals are, in general, difficult to compute. However, one can break up the action into two pieces, $H_0(x)$ and $H_1(x)$, and then by expanding in $H_1(x)$ and explicitly evaluating the functional integral, one may obtain a new (as far as the author can determine) series expansion for the transition amplitude coefficients. $H_0(x)$ acts as a free effective Hamiltonian, while $H_1(x)$ acts as a perturbing potential. In fact $H_1(x) = 0$ when the potential is constant. The leading term in this expansion is the same as the WKB approximation. This expansion, is however, not the same as semiclassical (WKB) or eikonal expansions since the former is in terms of $k'(x)/k(x)$ while the latter involves higher derivatives of the potential.

The second aspect of this paper is the use of functional integrals over anticommuting variables. Such objects were first introduced[1] to define Feynman integrals for theories with fermions. Anticommuting variables have also become popular with the advent of supersymmetry and graded Lie algebras.[2] The author believes they will become an extremely useful calculational tool in the future.[3] This paper demonstrates their power in a seemingly inapplicable problem. Most of the material on anticommuting variables is not new and thus presented in three appendices. The first shows how "continuous" multiplication of matrices can be put into a simple form using anticommuting functional integrals. The second develops a calculational technique used to obtain Eq. (II. 14) from Eq. (II. 12). It is a perturbative expansion in the off-diagonal elements of the "Hamiltonian" matrix, H. The third appendix derives the equations of motion. It also shows that even though $\eta(x)$ and $\eta^*(x)$ *completely* anticommute as "classical" variables, as operators they satisfy canonical anticommutation relations (CAR).

The method used to obtain the scattering formula, Eq. (II. 12), is simple. Take any localized potential and approximate it by a step function (see Figs. 1 and 2). The scattering due to this step function potential should be approximately the same as the true potential. The scattering due to a step function is easily computed. One uses the T matrix formalism[4] learned in first year quantum mechanics. To obtain the T matrix one need only multiply the T matrices for each step of the potential. If there are n steps this involves an n-fold multiplication of matrices. In the limit as $n \to \infty$ the step potential becomes the true potential, the approximation to the scattering becomes the exact scattering, and the n-fold multiplication of matrices becomes a "continuous" multiplication of matrices which is why anticommuting variables enter: They facilitate "continuous" multiplication of matrices.

One final remark: There appears to be no higher dimensional analog to Eq. (II. 12) because only in one dimension do T matrices multiply. This is unfortunate since most interesting scattering processes are not reducible to one dimension.

II. THE DERIVATION

Let V be a localized one-dimensional potential, i.e., $V(x) = V_R$ for $x \geq b$, $V(x) = V_L$ for $x \leq a$ with V_R, V_L constants. Also assume that V is a bounded C^2 function. Such a potential is shown in Fig. 1. The quantum mechanical scattering can be described in terms of the T matrix,

$$\begin{pmatrix} A_+(R) \\ A_-(R) \end{pmatrix} = \begin{pmatrix} T_{++} T_{+-} \\ T_{-+} T_{--} \end{pmatrix} \begin{pmatrix} A_+(L) \\ A_-(L) \end{pmatrix}. \tag{II. 1}$$

The letters "R" and "L" refer to "right" and "left." The numbers $A_+(R), A_-(R), A_+(L)$, and $A_-(L)$ are defined by Eqs. (II. 2) and (II. 3) as follows: Let $\psi(x)$ be the solution to the time independent Schrödinger equation,

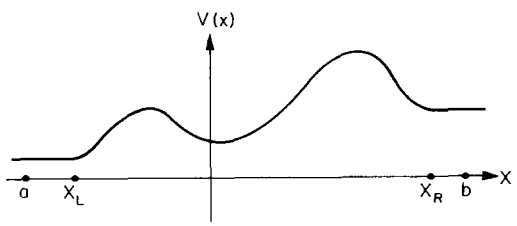


FIG. 1. A typical potential.

[a)]Research sponsored by the National Science Foundation under grant number PHY 74-08175-A02.

 0022-2488/78/1906-1438$1.00

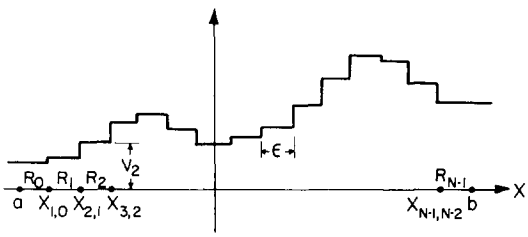


FIG. 2. A step function approximation. The potential of Fig. 1 has been approximated by a step function. The interval $[a, b]$ has been broken into N regions, R_i, in which the average value of the potential is V_i.

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) + V(x)\psi(x) = E\psi(x).$$

In the region to the right $(x > b)$

$$\psi(x) = A_+(R)\exp(ik_R \cdot x) + A_-(R)\exp(-ik_R \cdot x), \qquad \text{(II. 2)}$$

and in the region to left $(x < a)$

$$\psi(x) = A_+(L)\exp(ik_L \cdot x) + A_-(L)\exp(-ik_L \cdot x), \qquad \text{(II. 3)}$$

where

$$k_R \equiv \left(\frac{2m}{\hbar^2}(E - V_R)\right)^{1/2}, \qquad \text{(II. 4a)}$$

$$k_L \equiv \left(\frac{2m}{\hbar^2}(E - V_L)\right)^{1/2}, \qquad \text{(II. 4b)}$$

and in general

$$k(x) \equiv \left(\frac{2m}{\hbar^2}(E - V(x))\right)^{1/2}. \qquad \text{(II. 4c)}$$

For simplicity assume that the energy is greater than the maximum value of the potential. If this is not the case one must replace $k(x)$ by $\pm i\,k(x)$; the sign being determined by requiring exponential damping in the classically forbidden region.

The "+" and "−" signs refer to the sign of the phase of the wave. $\exp(ik \cdot x)$ is a wave moving to the right (and hence a "plus" wave), whereas $\exp(-ik \cdot x)$ is a wave moving to the left (and hence a "minus" wave) [if the time dependence of the wavefunction were made explicit, it would be $\exp(i(E/\hbar)t)$]. The T matrix spatially propagates from left to right the plus and minus waves. T_{++} represents the probability amplitude that a plus wave propagates through the potential and remains a plus wave while T_{-+} represents the probability amplitude that a plus wave propagates through the potential and becomes a minus wave. Similarly T_{+-} (and T_{--}) represents the probability of a minus wave propagating through the potential to become (or remain) a plus (minus) wave.

Let $R \equiv [a, b]$ be a region containing the potential. To solve this problem break R into a large number, N, of subregions, R_i, of width ϵ. $N\epsilon = b - a$. Replace V by an approximating step function, $V^{(N)}$ (see Fig. 2). Let V_i = the average value of V in R_i. $R_i \equiv [x_{i,i-1}, x_{i+1,i}]$ with $x_{i,i-1} \equiv a + i\epsilon$. One can take $V^{(N)} = \sum_i V_i \chi_{R_i}$, χ_{R_i} being the characteristic function for R_i. Define $k_i \equiv [(2m/\hbar^2) \times (E - V_i)]^{1/2}$, the wavenumber for the region, R_i.

The scattering due to V is approximately the same as that due to $V^{(N)}$ and in the limit as N goes to ∞ the approximation should become exact.

The T matrix for $V^{(N)}$ is simple,[4]

$$T^{(N)} = T(R, N)T(N, N-1)\cdots T(2, 1), \qquad \text{(II. 5)}$$

where $T(i, i-1)$ is the transition matrix for the single-step potential, $V(x) = V_i$ for $x > x_{i,i-1}$ and $V(x) = V_{i-1}$ for $x < x_{i,i-1}$,

$$T(i, i-1) = \begin{pmatrix} \frac{k_i + k_{i-1}}{2k_i}\exp[i(k_{i-1} - k_i)x_{i,i-1})] & \frac{k_i - k_{i-1}}{2k_i}\exp[-i(k_i + k_{i-1})x_{i,i-1}] \\ \frac{k_i - k_{i-1}}{2k_i}\exp[i(k_i + k_{i-1})x_{i,i-1}] & \frac{k_i + k_{i-1}}{2k_i}\exp[i(k_i - k_{i-1})x_{i,i-1}] \end{pmatrix}. \qquad \text{(II. 6)}$$

Since $k(x)$ is continuous, $k_{i+1} \approx k_i + \epsilon k'_{i+1}$, where k'_i is an appropriate approximation to $k'(x)$ in the regions R_i and R_{i+1}. To first order in ϵ

$$\frac{k_i - k_{i-1}}{2k_i} \to \frac{\epsilon k'_i}{2k_i}, \quad \frac{k_i + k_{i-1}}{2k_i} \to 1 - \frac{\epsilon}{2}\frac{k'_i}{k_i}, \quad \exp[\pm i(k_i - k_{i-1})x_{i,i-1}] \to 1 \pm i\epsilon k'_i x_{i,i-1},$$

$$\exp[\pm i(k_i + k_{i-1})x_{i,i-1}] \to \exp(\pm 2ik_1 x_{1,i-1}). \qquad \text{(II. 7)}$$

Rewriting

$$T(i, i-1) \approx 1 + \epsilon H(i, i-1) \approx \exp[\ln(1 + \epsilon H(i, i-1))] = \exp[\epsilon H(i, i-1)](1 + O(\epsilon)),$$

one obtains

$$H(i, i-1) = \begin{pmatrix} -\frac{k'_i}{2k_i} - ik'_i x_{i,i-1} & \frac{k'_i}{2k_i}\exp(-2ik_i x_{i,i-1}) \\ \frac{k'_i}{2k_i}\exp(2ik_i x_{i,i-1}) & -\frac{k'_i}{2k_i} + ik'_i x_{i,i-1} \end{pmatrix}. \qquad \text{(II. 8)}$$

Equation (II. 5) is replaced by

$$T = \exp[\epsilon H(R, N)]\exp[\epsilon H(N, N-1)]\cdots\exp[\epsilon H(2, 1)](1 + O(\epsilon)). \qquad \text{(II. 9)}$$

In Appendix A it is shown that

$$\lim_{N\to\infty}\{\exp[\epsilon H(R,N)]\exp[\epsilon H(N,N-1)]\cdots\exp[\epsilon H(2,1)]\}_{\alpha\beta}$$
$$=\int\int\mathcal{D}\eta\mathcal{D}\eta^*\eta(f)_\alpha\exp\{\int_a^b[\sum_\gamma\eta_\gamma'^*(x)\eta_\gamma(x)+\sum_{\gamma\zeta}\eta_\gamma^*(x)H_{\gamma\zeta}(x)\eta_\zeta(x)]\,dx\}\eta^*(0)_\beta. \tag{II.10}$$

Here $\eta'(x)=(d/dx)\eta(x)$ and $\int\int\mathcal{D}\eta\mathcal{D}\eta^*\sim\Pi_x\Pi_\alpha\, d\eta_\alpha(x)d\eta_\alpha^*(x)$. This is a functional integral over anticommuting variables $\eta(x)$, $\eta^*(x)$ with $H(x)$ an appropriate continuous version of $H(i,i-1)$,

$$H(x)=\begin{pmatrix}\dfrac{-k'(x)}{2k(x)}-ik'(x)x & \dfrac{k'(x)}{2k(x)}\exp[-2ik(x)x]\\ \dfrac{k'(x)}{2k(x)}\exp[2ik(x)x] & \dfrac{-k'(x)}{2k(x)}+ik'(x)x\end{pmatrix}. \tag{II.11}$$

Functional integrals over anticommuting variables accommodate continuous products of matrices. Note that Eq. (II. 9) is a product of factors which do not commute and hence does not exponentiate. Anticommuting variables allow one to exponentiate such a product recasting it into an "action" formalism.

The final solution, expressed in terms of a functional integral over anticommuting variables, is

$$T_{\alpha\beta}=\int\int\mathcal{D}\eta\mathcal{D}\eta^*(\eta(f))_\alpha\exp\{\int_a^b[\eta'^*(x)\eta(x)+\eta^*(x)H(x)\eta(x)]\,dx\}(\eta^*(0))_\beta. \tag{II.12}$$

This scattering problem has been reduced to solving the quantum mechanical fermion problem given by the Lagrangian, $\mathcal{L}(\eta^*,\eta)=\sum_\alpha\eta_\alpha^{*\prime}(x)\eta_\alpha(x)+\sum_{\alpha,\beta}\eta_\alpha^*(x)H_{\alpha\beta}(x)\eta_\beta(x)$. Since, in general, functional integrals are hard to solve, it is useful to obtain a perturbativelike expansion for Eq. (II. 12). Let $H=H_0+H_1$ with

$$H_0(x)=\begin{pmatrix}\dfrac{-k'(x)}{2k(x)}-ik'(x)x & 0\\ 0 & \dfrac{-k'(x)}{2k(x)}+ik'(x)x\end{pmatrix},\quad H_1(x)=\begin{pmatrix}0 & \dfrac{k'(x)}{2k(x)}\exp[-2ik(x)x]\\ \dfrac{k'(x)}{2k(x)}\exp[+2ik(x)x] & 0\end{pmatrix}, \tag{II.13}$$

and expand the exponential in powers of H_1. One can explicitly do the functional integrals term by term (done in Appendix B) obtaining a scattering formula with a simple physical interpretation,

$$T_{\alpha\beta}=G_0(b,a)_{\alpha\beta}+\sum_{n=1}^{\infty}\int\int\cdots\int_{b\geq y_n\geq\cdots\geq y_1\geq a}dy_1\cdots dy_n[G_0(b,y_n)H_1(y_n)G_0(y_n,y_{n-1})H_1(y_{n-1})\cdots G_0(y_2,y_1)H_1(y_1)G_0(y_1,a)]_{\alpha\beta}, \tag{II.14}$$

where

$$G_0(y_i,y_{i-1})=\left(\frac{k(y_{i-1})}{k(y_i)}\right)^{1/2}\times\begin{pmatrix}\exp[-iy_ik(y_i)]\exp\left[i\int_{y_{i-1}}^{y_i}k(x)\,dx\right]\exp[iy_{i-1}k(y_{i-1})] & 0\\ 0 & \exp[iy_ik(y_i)]\exp\left[-i\int_{y_{i-1}}^{y_i}k(x)\,dx\right]\exp[-iy_{i-1}k(y_{i-1})]\end{pmatrix}. \tag{II.15}$$

Note that because H_1 is an off diagonal matrix only even terms in n contribute to T_{++} and T_{--} while odd terms contribute to T_{+-} and T_{-+}. To see the physical significance of Eq. (II. 14) write out, for example, the first few terms of T_{++},

$$T_{++}=\left(\frac{k(a)}{k(b)}\right)^{1/2}\exp[-ibk(b)]\exp\left[i\int_a^b k(x)\,dx\right]\exp[iak(a)]+\int_a^b dy_2\int_a^{y_2}dy_1$$
$$\times\underbrace{\left[\left(\frac{k(y_2)}{k(b)}\right)^{1/2}\exp[-ibk(b)]\exp\left[i\int_{y_2}^b k(x)\,dx\right]\exp[iy_2k(y_2)]\right]}_{\text{(e)}}$$
$$\times\underbrace{\left[\frac{k'(y_2)}{2k(y_2)}\exp[-2iy_2k(y_2)]\right]}_{\text{(d)}}\underbrace{\left[\left(\frac{k(y_1)}{k(y_2)}\right)^{1/2}\exp[iy_2k(y_2)]\exp\left[-i\int_{y_1}^{y_2}k(x)\,dx\right]\exp[-iy_1k(y_1)]\right]}_{\text{(c)}}$$
$$\times\underbrace{\left[\frac{k'(y_1)}{2k(y_1)}\exp[2iy_1k(y_1)]\right]}_{\text{(b)}}\underbrace{\left[\left(\frac{k(a)}{k(y_1)}\right)^{1/2}\exp[-iy_1k(y_1)]\exp\left[i\int_a^{y_1}k(x)\,dx\right]\exp[iak(a)]\right]}_{\text{(a)}}+\cdots. \tag{II.16}$$

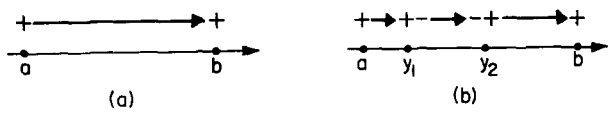


FIG. 3. Diagramatic expansion of T_{++}. Fig. 3(a) represents the probability amplitude that a plus wave propagates from a to b without phase flipping. This term is identical to the WKB approximation. Fig. 3(b) is the next leading contribution to T_{++}. A plus wave propagates from a to y_1. At y_1 the plus wave undergoes a phase flip and becomes a minus wave. The minus wave then propagates from y_1 to y_2. Again at y_2 a phase flip takes place, changing the minus wave into a plus wave. Finally, the plus wave propagates from y_2 to b.

The first term is the WKB approximation. I interpret it as the probability that a plus wave propagates from left to right through the potential without any phase flips [Fig. 3(a)]. The second term is a product of five factors each of which may be interpreted as follows: Factor (a) is the probability that a plus wave propagates forward from a to y_1. Factor (b) is the probability that the plus wave flips into a minus wave at y_1 and is called the plus to minus phase flip factor. Factor (c) is the probability that a minus wave moves forward from y_1 to y_2. Factor (d) is the minus to plus phase flip probability at y_2. Finally, factor (e), similar to factor (a), is the probability that a plus wave moves forward from y_2 to b. These five factors are schematically illustrated in Fig. 3(b). The phase flips may occur anywhere within the interval $[a, b]$ with the restriction that the minus to plus phase flip at y_2 occurs after the plus to minus phase flip at y_1. Thus one integrates over y_1 and y_2 with the restriction that $y_2 > y_1$.

The physical factors are summarized in Table I.

Once one has this physical interpretation of Eq. (II. 14) it's easy to write any term. For example, the lowest term in T_{+-} consists of the following three factors: One, the probability that a minus wave moves from a to y_1; two, the probability of a minus to plus phase flip at y_1; and three, the probability that a plus wave moves from y_1 to b,

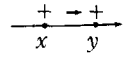
$$T_{+-} = \int_a^b dy G_{++}(b, y) F_{+-}(y) G_{--}(y, a) + \cdots . \qquad \text{(II. 17)}$$

One thus obtains the following scattering diagram rules:

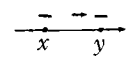
(i) To compute the Nth (N odd) contribution to T_{++} or T_{--}, partition the interval $[a, b]$ into N sections at the points $y_1 < y_2 < \cdots < y_{N-1}$.

(ii) Put in arrows between successive points and alternate plus and minus signs (as is illustrated in Fig. 3) with +'s at end points for T_{++} and −'s at end points for T_{--}.

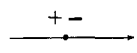
(iii) Put in a factor of $G_{++}(y, x)$ for



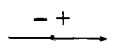
Put in a factor of $G_{--}(y, x)$ for



(iv) Put in a factor of $F_{-+}(y)$ for



Put in a factor of $F_{+-}(y)$ for



(v) Integrate over the partition points with the restriction $a \leq y_1 \leq y_2 \leq \cdots \leq y_{N-1} \leq b$.

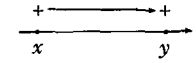
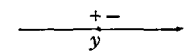
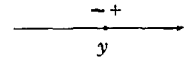
(vi) To compute the Nth (N even) contribution to T_{+-} or T_{-+} follow a similar procedure but partition $[a, b]$ into an even number of divisions.

Equation (II. 17) is the result for Fig. 4.

A minor point: If a and b are chosen to be bigger than necessary (i.e., $V'(x) \neq 0$ in a region strictly smaller than $[a, b]$), then the above formulas are still valid. The factors $\exp[-ibk(b)]$ and $\exp[iak(a)]$ appropriately compensate for the integration of $k(x)$ over regions where V is flat.

Note that although the leading term in T is the WKB

TABLE I. Physical interpretation.

Name	Physical interpretation	Diagram symbol	Formula
$G_{++}(y, x)$	probability that a plus wave at x propagates without phase flip to y		$\left(\frac{k(x)}{k(y)}\right)^{1/2} \exp[-iyk(y)] \times \exp[i\int_x^y k(z)\,dz] \exp[ixk(x)]$
$G_{--}(y, x)$	probability that a minus wave at x propagates to y without phase flip		$\left(\frac{k(x)}{k(y)}\right)^{1/2} \exp[iyk(y)] \times \exp[-i\int_x^y k(z)dz] \exp[-ixk(x)]$
$F_{-+}(y)$	plus to minus phase flip factor, the probability that a plus wave flips into a minus wave at y		$\frac{k'(y)}{2k(y)} \exp[2iyk(y)]$
$F_{+-}(y)$	minus to plus phase flip factor, the probability that a minus wave flips into a plus wave at y		$\frac{k'(y)}{2k(y)} \exp[-2iyk(y)]$

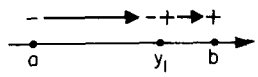


FIG. 4. Diagramatic interpretation of Eq. (II. 17). The $\xrightarrow[a]{-}\xrightarrow[y_1]{-}$ gives a factor of $G_{--}(y_1, a)$. $\xrightarrow[y_1]{-+}$ gives a factor of $F_{+-}(y)$. $\xrightarrow[y_1]{+}\xrightarrow[b]{+}$ gives a factor of $G_{++}(b, y_1)$. Finally, the phase flip may occur anywhere between a and b so that one must sum over all such locations. This introduces a factor of $\int_a^b dy_1$. One, thus, obtains Eq. (II. 7).

approximation, expansion Eq. (II. 14) is neither the WKB nor the eikonal expansion, as one can check. These latter expansions are in powers of $\hbar$ and involve higher derivative of k, whereas Eq. (II. 14) is a power series in (k'/k).

Many cancellations occur in the products of Eq. (II. 14). Let x_L and x_R be the exact end points V becomes flat, i.e., $V'(x) = 0$ for $x > x_R$, $V'(x) = 0$ for $x < x_L$. It is true that $k(a) = k(x_R)$ and $k(b) = k(x_L)$. Define

$$M_{++} = \left(\frac{k(b)}{k(a)}\right)^{1/2} \exp[ix_R k(x_R)] T_{++} \exp[-ix_L k(x_L)],$$

$$M_{+-} = \left(\frac{k(b)}{k(a)}\right)^{1/2} \exp[ix_R k(x_R)] T_{+-} \exp[ix_L k(x_L)],$$

$$M_{-+} = \left(\frac{k(b)}{k(a)}\right)^{1/2} \exp[-ix_R k(x_R)] T_{-+} \exp[-ix_L k(x_L)],$$

$$M_{--} = \left(\frac{k(b)}{k(a)}\right)^{1/2} \exp[-ix_R k(x_R)] T_{--} \exp[ix_L k(x_L)]. \tag{II. 18}$$

M has a simpler expansion,

$$M = M_0(x_R, x_L) + \sum_{n=1}^{\infty} \int\int \cdots \int_{x_L \le y_1 \le y_2 \le \cdots \le x_R} dy_1 \cdots dy_n \tag{II. 19}$$

$$\times M_0(x_R, y_n) U(y_n) M_0(y_n, y_{n-1}) U(y_{n-1}) \cdots M_0(y_1, x_L),$$

where

$$M_0(y_i, y_{i-1}) = \begin{pmatrix} \exp[i\int_{y_{i-1}}^{y_i} k(x)\,dx] & 0 \\ 0 & \exp[-i\int_{y_{i-1}}^{y_i} k(x)\,dx] \end{pmatrix}, \tag{II. 20}$$

$$U(y_i) = \begin{pmatrix} 0 & \frac{k'(y_i)}{2k(y_i)} \\ \frac{k'(y_i)}{2k(y_i)} & 0 \end{pmatrix},$$

which may be expressed in terms of a functional integral

$$M_{\alpha\beta} = \int\int \mathcal{D}\eta \mathcal{D}\eta^* (\eta(f))_\alpha \times \exp\left\{\int_{x_R}^{x_L} [\eta^{*\prime}(x)\eta(x) + \eta^*(x)\tilde{H}(x)\eta(x)]\,dx\right\} (\eta^*(0))_\beta, \tag{II. 21}$$

where

$$\tilde{H} = \begin{pmatrix} ik(x) & \frac{k'(x)}{2k(x)} \\ \frac{k'(x)}{2k(x)} & -ik(x) \end{pmatrix} = \tilde{H}_0 + U. \tag{II. 22}$$

APPENDIX A

Throughout these appendices one should think of x as a time variable and H as a Hamiltonian for a fermion system.

The purpose of this first appendix is to show how continuous matrix multiplication may be accommodated by functional integrals over anticommuting variables. Simple properties of anticommuting variables are discussed first. Latin letters will refer to space points whereas Greek letters will refer to matrix indices.

Let $H(x)$ be a smooth continuous $s \times s$ matrix function on $[a, b]$. Set $\Delta x = (b-a)/N$ and $x_j = j\Delta x + a$. Let $\eta_\alpha(x_j) \equiv \eta_\alpha(j)$ and $\eta_\alpha^*(x_j) \equiv \eta_\alpha^*(j)$ ($\alpha = 1, 2, \ldots, s$ and $j = 0, 1, 2, \ldots, N, f$) be a set of anticommuting variables:

$$\begin{aligned} &\{\eta_\alpha(i), \eta_\beta(j)\} = 0, \\ &\{\eta_\alpha(i), \eta_\beta^*(j)\} = 0, \\ &\{\eta_\alpha^*(i), \eta_\beta^*(j)\} = 0, \end{aligned} \tag{A1}$$

for all i, j, α, β.

Note that it is *not* true that $\{\eta_\alpha(i), \eta_\beta^*(j)\} = \delta_{ij}\delta_{\alpha\beta}$. This relation will be satisfied when inserted in a functional integral containing the action as will be shown in Appendix C.

Define integration[1] over anticommuting variables by

$$\int d\eta_\alpha(i) d\eta_\alpha^*(i) f = \begin{cases} 0 & \text{if } f \neq \eta_\alpha(i)\eta_\alpha^*(i), \\ 1 & \text{if } f = \eta_\alpha(i)\eta_\alpha^*(i). \end{cases} \tag{A2}$$

In Eq. (A2) neither α nor i are summed over. At each "x value," $j\epsilon + a$, define $\int d\eta(j) d\eta^*(j) = \Pi_\alpha \int d\eta_\alpha(j)\, d\eta_\alpha^*(j)$.

Consider

$$T_{\alpha\beta}^{(N)} \equiv \int d\eta(f)\, d\eta^*(f) \int d\eta(N)\, d\eta^*(N) \cdots \int d\eta(0) d\eta^*(0) \times \eta_\alpha(f) \exp\left(\sum_{j=0}^{N} A_j\right) \eta_\beta^*(0), \tag{A3}$$

where the A_j which make up the "action" are

$$A_j \equiv A(x_j) \equiv \sum_\alpha [\eta_\alpha^*(j+1) - \eta_\alpha^*(j)]\eta_\alpha(j) + \sum_{\alpha,\beta} \eta_\alpha^*(j+1) H(x_j)_{\alpha\beta} \eta_\beta(j) \Delta x. \tag{A4}$$

Note that

$$\int d\eta(j) d\eta^*(j) \exp[\sum_\alpha \eta_\alpha(j)\eta_\alpha^*(j)] \times \begin{Bmatrix} 1 \\ \eta_\gamma(j)\eta_\gamma^*(j) \\ \eta_{\gamma_1}(j)\eta_{\gamma_1}^*(j)\eta_{\gamma_2}(j)\eta_{\gamma_2}^*(j) \quad (\gamma_1 \neq \gamma_2) \\ \vdots \\ \eta_1(j)\eta_1^*(j) \cdots \eta_{\gamma-1}(j)\eta_{\gamma-1}^*(j)\eta_{\gamma+1}(j)\eta_{\gamma+1}^*(j) \cdots \eta_s(j)\eta_s^*(j) \\ \eta_1(j)\eta_1^*(j)\eta_2(j)\eta_2^*(j) \cdots \eta_s(j)\eta_s^*(j) \end{Bmatrix} = 1. \tag{A5}$$

All other integrals are zero.

Expand the $\sum_{j=0}^{N}[\eta^*(j+1)\eta(j)+\eta^*(j+1)H(j)\eta(j)\Delta x]$ part of the exponent in Eq. (A3) leaving $\exp[\sum_j \eta(j)\eta^*(j)]$. Using Eq. (A5) the integrations are easy to do. First do the $\eta(0),\eta^*(0)$ integrals. The relevant factor is

$$\int d\eta(0)d\eta^*(0)\exp[\eta(0)\eta^*(0)]\sum_n \frac{1}{n!}[\eta^*(1)\eta(0) + \Delta x\eta^*(1)H(x_1)\eta(0)]^n\eta_\beta^*(0) = \eta_\gamma^*(1)[I+\Delta xH(x_1)]_{\gamma\beta}. \tag{A6}$$

The last step follows since a single $\eta_\alpha^*(0)$ being present implies that only the $n=1$ term contributes. I in Eq. (A6) is the identity matrix.

The $\eta(0)$, $\eta^*(0)$ integrations have left the remaining integrals in exactly the same form. After doing $\eta_1, \eta_1^* \cdots \eta_N, \eta_N^*$ integrations in that order one obtains

$$T_{\alpha\beta}^{(N)} = \int d\eta(f)\,d\eta^*(f)\exp[\eta(f)\eta^*(f)]\eta_\alpha(f) \times \eta_\gamma^*(f)\{[I+\Delta xH(N)][I+\Delta xH(N-1)]\cdots[I+\Delta xH(1)]\}_{\gamma\beta} = \{[I+\Delta xH(N)]\cdots[I+\Delta xH(1)]\}_{\alpha\beta}. \tag{A7}$$

Because

$$T^{(N)} = \exp[\Delta xH(N)]\exp[\Delta xH(N-1)]\cdots \times \exp[\Delta xH(1)][I+O(\Delta x)], \tag{A8}$$

$$T_{\alpha\beta} = \lim_{N\to\infty} T_{\alpha\beta}^{(N)} = \int \mathcal{D}\eta(x)\mathcal{D}\eta^*(x)\,\eta_\alpha(f)e^A\eta_\beta^*(0), \tag{A9}$$

with A the continuous version of $\sum_i A_i$,

$$A = \int_a^b [\eta'^*(x^+)\eta(x) + \eta^*(x^+)H(x)\eta(x)]\,dx. \tag{A10}$$

The "$'$" denotes derivative with respect to x and the "$+$" denotes a slightly later position ($x^+ = x+\epsilon$) to remind one that η'^* and η are not multiplied at the same space point.

Equation (A8) is the result Eq. (II. 10) used in Sec. II.

APPENDIX B

This Appendix proves Eq. (II. 14) from Eqs. (II. 12) and (II. 13), by writing $H = H_0 + H_1$ and expanding the exponent including H_1:

$$T_{\alpha\beta} = \int\int \mathcal{D}\eta\mathcal{D}\eta^*\eta_\alpha(f)\exp\left[\int_a^b \eta'^*(x)\eta(x) + \eta^*(x)H_0(x)\eta(x)\,dx\right] \times\sum_{n=0}^{\infty}\frac{1}{n!}\left[\int_a^b dx\eta^*(x)H_1(x)\eta(x)\right]^n\eta_\beta^*(0)$$
$$= \int\int \mathcal{D}\eta\mathcal{D}\eta^*\eta_\alpha(f)\exp\left\{\int_a^b [\eta'^*(x)\eta(x)+\eta^*(x)H_0(x)\eta(x)]\,dx\right\} \times\left(\sum_{n=0}^{\infty}\int_{a\le y_1\le\cdots\le y_n\le b} dy_1\cdots\int dy_n\,\eta^*(y_n)H_1(y_n)\eta(y_n)\cdots \times\eta^*(y_1)H_1(y_1)\eta(y_1)\right)\eta_\alpha^*(0). \tag{B1}$$

The last step follows since the integrand in $[\int_a^b dx\eta^*(x) \times H_1(x)\eta(x)]^N = \int_a^b dy_1\int_a^b dy_2\cdots\int_a^b dy_N\,\eta^*(y_1)H_1(y_1)\eta(y_1)\cdots \times\eta^*(y_N)H_1(y_N)\eta(y_N)$ is a completely symmetric function of $y_1, y_2, \ldots, y_N$ because the $\eta^*\eta$ products commute.

The $n=0$ term is easy to evaluate. It is the continuous matrix product, $T^0 = \Pi_i \exp[\epsilon H_0(i)]$, and because H_0 is diagonal this product reduces to ordinary integrals,

$$T^0 = \begin{pmatrix} \exp[\int (H_0)_{++}dx] & 0 \\ 0 & \exp[+\int (H_0)_{--}dx] \end{pmatrix} = \left(\frac{k(b)}{k(a)}\right)^{1/2}\begin{pmatrix} \exp[-ik(b)b]\exp[i\int_a^b k(x)\,dx]\exp[ik(a)a] & 0 \\ 0 & \exp[ik(b)b]\exp[-i\int_a^b k(x)\,dx]\exp[-ik(a)a] \end{pmatrix}. \tag{B2}$$

An integration by parts has been done in obtaining Eq. (B2).

T^0 is $G^0(b,a)$. So the zeroth order term in Eq. (II. 14) agrees with the zeroth order term in Eq. (B1). To evaluate the higher order terms, latticize the functional integral in Eq. (B1). Doing the integration variables up to y_1 will yield a factor of

$$\int d\eta(y_1)d\eta^*(y_1)\exp[\eta(y_1)\eta^*(y_1)+\eta^*(y_1+\epsilon)\eta(y_1) + \eta^*(y_1+\epsilon)H_0\eta(y_1)\epsilon]\eta^*(y_1+\epsilon)H_1(y_1)\eta(y_1)\eta_\gamma^*(y_1)G_{\gamma\beta}^0(y_1,a) = \eta_\gamma^*(y_1+\epsilon)[H_1(y_1)G^0(y_1,a)]_{\gamma\beta} \tag{B3}$$

and when the rest of the anticommuting integrations are done one obtains

$$G^0(b,y_N)H_1(y_N)G^0(y_N,y_{N-1})\cdots H_1(y_1)G^0(y_1,a) \tag{B4}$$

as the integrand to the y_j integrals thus obtaining Eq. (II. 14).

APPENDIX C

This Appendix derives the equations of motion and canonical anticommutation relations in a manner similar to the boson case.[5] They are easily obtainable by using anticommuting derivatives $d/d\eta$ and $d/d\eta^*$. Anticommuting derivatives are similar to ordinary derivatives except that one must be careful to include a minus sign when $d/d\eta$ or $d/d\eta^*$ moves past an anticommuting variable. For example $(d/d\eta_1)(\eta_1\eta_2^*) = \eta_2^*$ but $(d/d\eta_1)\eta_2^*\eta_1 = -\eta_2^*(d/d\eta_1)\eta_1 = -\eta_2^*$. This is consistent since $(d/d\eta_1)\times(\eta_2^*\eta_1) = (d/d\eta_1)(-\eta_1\eta_2^*) = -\eta_2^*$. For the details in anticommuting derivatives see Ref. 1.

One can integrate by parts in an anticommuting vari-

able integral. Let $F(\eta, \eta^*)$ and $G(\eta, \eta^*)$ be two arbitrary functions in the η and η^* variables. The rules[1] are

$$\int d\eta\, d\eta^* F \frac{\vec{d}}{d\eta} G = \int d\eta\, d\eta^* F \frac{\overleftarrow{d}}{d\eta} G, \qquad \text{(C1)}$$
$$\int d\eta\, d\eta^* F \frac{\vec{d}}{d\eta^*} G = \int d\eta\, d\eta^* F \frac{\vec{d}}{d\eta^*} G.$$

The simplest example is $1 = \int d\eta\, d\eta^* \eta\eta^* (d/d\eta)\eta = \int d\eta\, d\eta^* \eta\eta^* (\vec{d}/d\eta)\eta = -\int d\eta\, d\eta^* \eta^* \eta = \int d\eta\, d\eta^* \eta\eta^* \stackrel{\checkmark}{=} 1$. Applying this to functional integrals,

$$\int\int \mathcal{D}\eta \mathcal{D}\eta^* e^A \frac{\vec{d}}{d\eta_\gamma(y)} F = \int\int \mathcal{D}\eta \mathcal{D}\eta^* e^A \frac{\overleftarrow{d}}{d\eta_\gamma(y)} F, \qquad \text{(C2)}$$
$$\int\int \mathcal{D}\eta \mathcal{D}\eta^* e^A \frac{\vec{d}}{d\eta_\gamma^*(y)} F = \int\int \mathcal{D}\eta \mathcal{D}\eta^* e^A \frac{\overleftarrow{d}}{d\eta_\gamma^*(y)} F,$$

where A is given by Eq. (A10).

Choose $F = \psi_f[\eta_\alpha(f), \eta_\alpha^*(f)]\psi_0[\eta_\alpha(0), \eta_\alpha^*(0)]$ with ψ_f and ψ_0 arbitrary and y to be in (a, b). Latticize the functional integral. One obtains using Eq. (C2)

$$0 = \langle \eta_\gamma^*(y+\epsilon) - \eta_\gamma^*(y) + \epsilon\eta_\delta^*(y+\epsilon) H_{\delta\gamma}(y)\rangle, \qquad \text{(C3)}$$
$$0 = \langle \eta_\gamma(y) - \eta_\gamma(y-\epsilon) - \epsilon H_{\gamma\delta}(y-\epsilon)\eta_\delta(y)\rangle,$$

where $\langle f\rangle$ means $\int\int \mathcal{D}\eta \mathcal{D}\eta^* e^A f \psi_f \psi_0$. The continuous versions of (C3) are

$$\left\langle \frac{d\eta^*}{dy} + \eta^* H \right\rangle = 0, \quad \left\langle \frac{d\eta}{dy} - H\eta \right\rangle = 0, \qquad \text{(C4)}$$

which one might guess by a variational calculation of the action. Equations (C4) are operator equations, satisfied no matter what "initial and final states" are used. Hence, $d\eta^*/dy + \eta^* H = d\eta/dy - H\eta = 0$ as operator equations.

Choosing $F = \eta_\mu(y)\psi_f\psi_0$ and $F = \eta_\mu^*(y)\psi_f\psi_0$ one obtains

$$\langle \delta_{\gamma\mu}\rangle = \langle [\eta_\gamma^*(y+\epsilon) - \eta_\gamma^*(y) + \epsilon\eta_\lambda^*(y+\epsilon) H_{\lambda\gamma}(y)]\eta_\mu(y)\rangle,$$
$$\langle \delta_{\gamma\mu}\rangle = \langle [\eta_\gamma(y) - \eta_\gamma(y-\epsilon) - \epsilon H_{\gamma\lambda}(y)\eta_\lambda(y-\epsilon)]\eta_\mu^*(y)\rangle,$$
$$0 = \langle [\eta_\gamma^*(y+\epsilon) - \eta_\gamma^*(y) + \epsilon\eta_\lambda^*(y+\epsilon) H_{\lambda\gamma}(y)]\eta_\mu^*(y)\rangle, \qquad \text{(C5)}$$
$$0 = \langle [\eta_\gamma(y) - \eta_\gamma(y-\epsilon) - \epsilon H_{\gamma\lambda}(y)\eta_\lambda(y-\epsilon)]\eta_\mu(y)\rangle,$$

which are as $\epsilon \to 0$ the operator equations

$$\delta_{\gamma\mu} = \eta_\gamma^*(y^+)\eta_\mu(y) + \eta_\mu(y)\eta_\gamma^*(y) = \eta_\mu^*(y)\eta_\gamma(y^-) + \eta_\gamma(y)\eta_\mu^*(y), \qquad \text{(C6)}$$
$$0 = \eta_\gamma^*(y^+)\eta_\mu^*(y) + \eta_\mu^*(y)\eta_\gamma^*(y) = \eta_\mu(y)\eta_\gamma(y^-) + \eta_\gamma(y)\eta_\mu(y).$$

Thus as operators the anticommuting variables satisfy canonical anticommutation relations.

ACKNOWLEDGMENT

I would like to thank K. Bardakci for reading a preliminary draft of this paper and encouraging its completion.

[1] D. J. Candlin, Nuovo Cimento **4**, 231 (1956); F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966); J. Schwinger, Phys. Rev. **82**, 914 (1951); J. Schwinger, Proc. Natl. Acad. Sci. **47**, 1075 (1961).
[2] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[3] A recent interesting paper which develops and uses anticommuting variables in field theory is M. B. Halpern, A. Jevicki, and P. Senjanovic, Phys. Rev. D **16**, 2476 (1977).
[4] E. Merzbacher, *Quantum Mechanics* (Wiley, New York 1970), see Chap. 6. Merzbacher denotes the T matrix by M.
[5] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965). See Chaper 7.
[6] J. Schwinger, Phys. Rev. **82**, 914 (1951).
[7] Proc. Natl. Acad. Sci. **47**, 1075 (1961).

A Chebyshev polynomial method for computing analytic solutions to eigenvalue problems with application to the anharmonic oscillator

John P. Boyd[a)]

Advanced Study Program, National Center for Atmospheric Research, [b)] *Boulder, Colorado 80303*
(Received 4 October 1976)

In this work, I present a two-part algorithm which employs Chebyshev polynomials to compute the eigenvalues and eigenfunctions of an ordinary differential equation and to represent their dependence on a problem parameter λ. To illustrate these methods, I apply them to the anharmonic oscillator where the parameter λ is the coupling constant. I show that these techniques are simple, efficient, and easy to program for a computer. Unlike perturbation methods or numerical tables, the Chebyshev algorithms calculate explicit, highly uniform polynomial approximations to each function on any interval on which the function is analytic. Perhaps the most important result of this work is that although the two Chebyshev algorithms are *numerical*, all the results are *analytic*—a dramatic counterexample to the conventional wisdom which holds that to obtain nonnumerical solutions, one must use analytic, paper-and-pencil techniques. Thus, the two-part Chebyshev approach, despite its numerical basis, should be regarded principally as a new and powerful method for computing *analytic* solutions which will succeed where ordinary (and even extraordinary) perturbation methods fail.

1. INTRODUCTION

In this work, I describe a two-part algorithm which uses Chebyshev polynomials to compute efficient, highly uniform representations for the dependence of the eigenvalues and eigenfunctions of an ordinary differential equation upon a parameter λ. No claim is made that either individual part is original; it is the use of these two algorithms in combination that is new. For the anharmonic oscillator, λ is the coupling constant which measures the strength of the x^4 term which is added to the potential of Schrödinger's equation for the ordinary harmonic oscillator. Traditionally, parametric dependence has been studied in two ways: perturbation theory and numerical tables. The former is restricted to a limited range of the parameter. Numerical tables do not provide an explicit representation of the dependence of various quantities on λ and interpolation between tabulated values will only yield one or two place accuracy unless the tables are very large, which is both unwieldy and computationally expensive. In contrast, the Chebyshev method produces explicit polynomial approximations for the dependence of each quantity on λ. Unlike perturbation series, the Chebyshev formulas can be applied to any interval on which the function of λ is analytic, regardless of the location of off-interval singularities, and the approximations are highly *uniform* over the whole interval.

Besides these advantages, the Chebyshev methods have other virtues, and the greatest of them is generality. Neither part of the algorithm employs special tricks, so the only problem-dependent parameters are the differential equation itself and the domain in x—λ space. The pseudospectral eigenvalue routine can usually compute enough modes to high accuracy to overlap with high precision WKB calculations, which become better as the mode number increases. Because of the rapid convergence of Chebyshev series, it is possible to represent the eigenfunctions as double Chebyshev series in λ and x over a range where other representations, such as double Taylor series, would either diverge or require too many terms to be useful. As an added bonus, we can differentiate the Chebyshev series with respect to λ and evaluate at $\lambda = 0$ (or any other value of λ) to check analytically computed perturbation series. Both parts of the algorithm are easy to program: Excluding comment cards and "canned" EISPACK routines, but including rather elaborate input and output statements, all the routines for the first part required 200 FORTRAN statements and all for the second, including the calculation of Taylor series, needed 240. Finally, if we calculate enough terms in the Chebyshev series to obtain more accuracy than we need in our final results, we can use the properties of Chebyshev polynomials to bound the error in the uniform norm for truncation after a given number of terms. As I describe with examples, we can determine, almost by inspection, how many Chebyshev polynomials we should retain to keep the error below a specified tolerance over the whole interval.

The first part of the algorithm consists of solving for the eigenfunctions and eigenvalues for a discrete set of values of λ. I employed the Chebyshev pseudospectral technique described in the next section. This method is accurate, economical, and easy to program, and usually works well even when a different expansion basis might seem more logical. To emphasize this, I have chosen to illustrate the algorithm by applying it to the anharmonic oscillator where Hermite functions (the exact eigenfunctions for $\lambda = 0$) are the natural basis. However, the second part will apply without change to eigenvalues and eigenfunctions computed in any way. Wavefunctions in quantum chemistry often depend nonlinearly on a number of dependent variational param-

[a)] Present address: Department of Atmospheric and Oceanic Sciences, University of Michigan, Ann Arbor, Michigan 48109.

[b)] The National Center for Atmospheric Research is sponsored by the National Science Foundation.

 0022-2488/78/1906-1445$1.00

eters (see Kauzman[1] or any standard text for examples), but these parameters can be treated exactly like the coefficients of a series expansion and be represented as polynomials in λ by the techniques I will describe below.

The second part of the algorithm converts the M-dimensional vector of the values of a function q at the discrete set of points λ_i, $i=1,\ldots,M$, into first a truncated Chebyshev series in λ and then into an ordinary polynomial by two successive matrix multiplications, where q is either an eigenvalue or a particular expansion coefficient. (Details are given in the next section.) The coefficients of the Chebyshev expansion are given rigorously by integrals defined by Eq. (2.2).

The discrete algorithm is equivalent to evaluating these integrals by M-point Gaussian quadrature. As I explain below, the errors introduced by the quadrature are no larger than those of the M-term polynomial approximation itself. The two square matrices that multiply the vector of values of q are given by explicit formulas.

In Sec. 2, I describe these algorithms in much greater detail and apply them to the anharmonic oscillator in Sec. 3. Good parametrizations for the eigenvalues already exist thanks to Hioe *et al.*,[2] and to emphasize that my primary purpose is to illustrate the Chebyshev methods rather than to study exhaustively the anharmonic oscillator, I have presented results chiefly for the ground and second excited states after showing the techniques are applicable to at least the lowest forty modes in principle. (I show that we can accurately compute expansions for at least forty modes: Convergence in x is *slower* for the higher order modes.) Two simple quadratic polynomials, one for small λ, the other for large, give the ground state energy to within one part in 700 for all positive values of λ—an order of magnitude improvement over comparable or more complicated formulas of Hioe *et al.*[2] I also present wholly new results: two polynomials in x and λ, both of fifteen terms, which give the ground state eigenfunction to within a maximum error of 1% of the peak amplitude of the mode for all positive λ and all x, and similarly for the second excited state with polynomials of 24 terms. I make a few closing remarks in Sec. 4 and discuss the theoretical and empirical convergence of Chebyshev series for functions which have asymptotic expansions about one endpoint (background material needed for the example) in Appendix B.

2. ALGORITHMS

Before describing the algorithms themselves, I will briefly discuss the properties of Chebyshev polynomials which make these algorithms effective. Least squares polynomial approximation is a standard topic of every first course in numerical analysis. The unique polynomial $P_N(x)$ which minimizes the weighted least square error on $[a,b]$ with the weighting function $r(x)$

$$\int_a^b [f(x)-P_N(x)]^2 r(x)dx$$

is given by

$$P_N(x)=\sum_{n=0}^{N} a_n\theta_n(x), \tag{2.1}$$

$$a_n=C_n^{-1}\int_a^b f(x)\theta_n(x)r(x)dx, \tag{2.2}$$

where the set of polynomials $\{\theta_n(x)\}$, each of degree equal to its subscript, are orthogonal with respect to the inner product

$$(\theta_n,\theta_m)\equiv\int_a^b \theta_n(x)\theta_m(x)r(x)dx \tag{2.3}$$

$$=\begin{cases}0 & \text{if } m\neq n,\\ C_n & \text{if } m=n.\end{cases} \tag{2.4}$$

For a given interval and weight function, the polynomials $\{\theta_n\}$ can always be constructed from the powers of x: $\{1,x,x^2,\cdots\}$ by the Gram–Schmidt orthogonalization process, which is discussed in Young and Gregory's[3] text along with all aspects of least squares approximation. The Chebyshev polynomials correspond to the special case $a=-1$, $b=1$, and $r(x)=(1-x^2)^{-1/2}$. The C_n are given in (A11).

What sets the Chebyshev polynomials apart from their peers is a simple argument given in Fox and Parker,[4] which shows that, for a general infinitely differentiable function, Chebyshev polynomials converge faster than any other member of the class of Jacobi polynomials, which includes Legendre polynomials as another special case and ordinary Taylor expansions as a limiting case. Like all members of this class, Chebyshev polynomials have the property of *infinite order convergence*: Asymptotically, the error decreases faster than any finite power of n. Furthermore, Chebyshev expansions converge inside the largest ellipse with foci at $(-1,1)$ within which the function is analytic. (See Appendix B for details.) Thus, unlike ordinary Taylor expansion, Chebyshev expansions will always converge on *any* interval on which the function is analytic.

The Chebyshev polynomials also have excellent properties with respect to the so-called uniform or Chebyshev norm which is defined by

$$L_\infty \equiv \max_{x\in[a,b]} |f(x)-P_n(x)|. \tag{2.5}$$

It has long been known (see Todd[5] for proof) that if $f(x)$ is analytic on $[a,b]$, then there exists a unique polynomial of degree n or less that minimizes the error in the uniform norm over the class of polynomials of degree n or less. Furthermore, if we denote the error in the uniform norm by δ, then the pointwise error will be exactly $\pm\delta$ at $n+2$ points on the interval for the best approximating polynomial. Unfortunately, no finite algorithm for computing this best polynomial approximation in the uniform norm is known, but reliable iterative procedures for computing the best polynomial on a discrete set of points in a finite number of steps have been developed: the Remes algorithm (see Young and Gregory[3] for a discussion).

A truncated Chebyshev series is generally not equal

to the best polynomial approximation of the same degree in the uniform norm, but empirically, it is usually quite close. Theoretically, Powell[6] proved that the error in the uniform norm of a Chebyshev series truncated after $N+1$ terms and computed by interpolation on the $N+1$ points given by (2.14) below, call it S_N, is related to the error of the best approximating polynomial in the uniform norm of the same degree, call its error E_N, by the inequality $S_N < (1+\phi_N)E_N$, where $\phi_N < 2.5$ for $N<10$ and $\phi_N < 5.5$ for $N<1000$. Since, once the degree of the truncated Chebyshev series is high enough to give a moderately good approximation, adding just one more term will usually decrease the pointwise error by at least a factor of three, it is rarely worthwhile to improve upon a truncated Chebyshev series to reduce the error in the uniform norm: It is far easier to simply add one more term. Even if we wished to compute the best polynomial approximation of a given degree n, the Remes algorithm demands a means of evaluating $f(x)$ for many values of x and also a first guess. A Chebyshev expansion of high degree and the truncated Chebyshev expansion of lower degree n, respectively, are good means for meeting these needs of Remes' procedure.

No approximation is useful without some measure of accuracy, but the Chebyshev polynomials satisfy the simple bound

$$|T_n(x)| \leq 1 \quad \text{for all } x\in[-1,1] \text{ and for all } n \tag{2.6}$$

since the polynomials are defined by

$$T_n(\cos\theta) \equiv \cos n\theta \quad \text{for } \theta\in[0,\pi]. \tag{2.7}$$

This means that we can bound the error in the *uniform* norm for a truncated Chebyshev series merely by summing the absolute values of the coefficients of all the neglected terms. When we have computed only a finite number of coefficients and these values are contaminated by numerical roundoff and quadrature errors, we cannot compute this error bound sum exactly. However, because of the property of infinite order convergence described above, coefficients usually diminish rapidly after some n which is the minimum necessary to achieve moderate accuracy, so if we compute a number of terms above this minimum n, neglect of higher order uncalculated coefficients will produce only a tiny error in calculating error bounds. Thus, unless we want a Chebyshev approximation whose error is only slightly larger than that of the computer precision we are employing, (2.6) is a powerful and reliable tool for estimating the maximum error in the Chebyshev norm. Furthermore, (2.6) is a *tight* bound on the values of the Chebyshev polynomials for all n. Because of this and the rapid convergence of Chebyshev series, which usually means that the first neglected term is not significantly cancelled by the second neglected term to make the error smaller than the magnitude of the first neglected term alone, error bounds on a series computed by using (2.6) are usually fairly tight: The true error in the Chebyshev norm is normally only a little smaller than the sum of the absolute values of the neglected coefficients.

Over-all, therefore, Chebyshev polynomials are about as efficient and accurate as any form of polynomial approximation. Because of this, I chose to use them not only to parametrize the dependence of the eigenvalues and eigenfunctions on the parameter λ but also to compute the eigenvalues and eigenfunctions themselves. To simplify programming, all the algorithms below automatically transfer all calculations via a simple linear transformation from $[a,b]$ to the standard interval $[-1,1]$. The truncated Chebyshev series and ordinary polynomials whose coefficients are given in the tables in the next section are always evaluated with an argument that lies on $[-1,1]$.

The simplest way to insure that the calculated eigenfunctions satisfy the appropriate homogeneous boundary conditions at the end points is to rearrange the Chebyshev polynomials into a new set of basis functions that individually satisfy the boundary conditions. For $u=0$ at a,b, we can use

$$\theta_n(x) = \begin{cases} T_n(x) - T_0(x), & n \text{ even}, \\ T_n(x) - T_1(x), & n \text{ odd}. \end{cases} \tag{2.8}$$

When the equation is symmetric about the midpoint of the interval, as is true of the anharmonic oscillator, we only need to compute on the half interval $[0, b]$. We can use (2.8) to compute the antisymmetric modes, but the symmetric modes must satisfy $du/dx=0$ at the midpoint of the interval, so we can use the basis

$$\begin{aligned} \theta_0(x) &= -\tfrac{1}{3}T_2(x) + \tfrac{4}{3}T_1(x) + \tfrac{5}{3}T_0(x), \\ \theta_n(x) &= -[n^2/(n+2)^2]T_{n+2}(x) + T_n(x) + \{[n^2/(n+2)^2] \\ &\quad - 1\}T_0. \end{aligned} \tag{2.9}$$

(There is a good deal of freedom in choosing new basis sets: Results will be little affected by the particular choice one makes unless one defines basis functions which are almost linearly dependent, which can lead to numerical ill-conditioning and severe errors.)

We now assume

$$u(x) = \sum_{n=0}^{N-1} u_n\theta_n(y), \tag{2.10}$$

$$y \equiv 2[x-(a+b)/2]/(b-a) \tag{2.11}$$

and substitute this in the differential equation (primes denote differentiation with respect to x)

$$a_2(x)u'' + a_1(x)u' + a_0(x)u = E(\lambda)q(x)u. \tag{2.12}$$

There are two popular ways of obtaining N algebraic equations for the expansion coefficients u_n. Galerkin's method, also called the spectral method, involves multiplying (2.12), after substitution of (2.10), by each of the first N basis functions in turn and integrating over the interval with the appropriate weight function, which is $(1-x^2)^{-1/2}$ for Chebyshev polynomials, and equating the result to 0. The pseudospectral or collocation method that I will use consists of demanding that the truncated expansion satisfy (2.12) exactly at N colloca-

tion points:

$$\sum_{n=0}^{N-1} u_n[a_2(x_i)\theta_n''(y_i)/d^2 + a_1(x_i)\theta_n'(y_i)/d + a_0(x_i)\theta_n(y_i)]$$
$$= E(\lambda)\sum_{n=0}^{N-1} q(x_i)\theta_n(y_i)u_n \quad \text{for } i=1,2,\ldots,N \tag{2.13}$$

(primes denote differentiation with respect to y),

$$y_i = -\cos[(2i-1)\pi/2N], \tag{2.14}$$

$$x_i = 0.5(a+b) + 0.5(b-a)y_i, \tag{2.15}$$

$$d = 0.5(b-a). \tag{2.16}$$

Equation (2.13) actually describes a generalized eigenvalue problem, i.e., the matrix multiplying E is not the identity but rather a general matrix M, but we can easily remedy this by multiplying through by M^{-1}.

Superficially, the spectral and pseudospectral methods are quite different, but the collocation points given by (2.14) are actually the Gaussian quadrature abscissae for the Chebyshev interval and weight function. Multiplying each row of (2.13) by $\theta_m(y_i)$ and summing over i for $m=0,1,..,N-1$, we see that the pseudospectral method is identical with Galerkin's method if we evaluate the integrals of the latter by N-point Gaussian quadrature.

If the exact Chebyshev coefficients of a function $f(x)$ are denoted by f_n, then the coefficients computed by N-point Gaussian quadrature will be (Fox and Parker[4])

$$f_n^G = f_n - (f_{2N-n}+f_{2N+n}) + (f_{4N-n}+f_{4N+n}) - \cdots. \tag{2.17}$$

Thus, if the errors from Gaussian quadrature are unacceptably large, the errors in even the exact Chebyshev series will also be unacceptably larger after N-term truncation. The errors in N-point quadrature and the N-term Chebyshev series are roughly the same. There is usually little point, therefore, in evaluating the spectral integrals by a quadrature of higher degree; one can usually obtain a greater error reduction by increasing N by one even if N is large.

Like other orthogonal polynomials, the Chebyshev polynomials can be efficiently evaluated by a three-term recursion relation, and its derivatives can be evaluated by similar three-term recursion relations that follow from differentiating that for $T_n(x)$ itself. Truncated Chebyshev series of N terms can be summed in $O(N)$ operations by an equally simple recursive formula given by Fox and Parker.[4] These expressions are given in the appendices.

Even if one wants to represent the eigenfunctions in terms of a non-Chebyshev basis—spherical harmonics or Hermite functions, for example—it may still be convenient to do the eigenfunction calculations via Chebyshev polynomials. Thanks to the Chebyshev sum formula, the work of evaluating the projection integrals similar to (2.2) to convert all the computed eigenfunctions into a different basis will be small compared to that needed to compute the eigenfunctions in the first place if any efficient quadrature scheme is used. However, it is usually easy to rewrite the whole psuedospectral algorithm in terms of other orthogonal polynomials, and so we can take our choice.

The final step is to call a standard "black box" subroutine to solve the algebraic eigenproblem posed by (2.13). For the calculations I present in the next section, I employed the QR algorithm from the EISPACK library of subroutines distributed by Argonne National Laboratory. Empirically, the QR algorithm requires $O(15\ N^3)$ operations to find all the eigenvalues and eigenfunctions (Stephen Orszag, unpublished *M.I.T.* lecture notes), and so, for computational efficiency, it is extremely important to minimize the number of degrees of freedom N. Because of this as discussed by Orszag[7] and Boyd,[8] Galerkin methods are usually much superior to finite difference methods for converting a differential equation eigenvalue problem into an algebraic characteristic value problem except for ordinary (as opposed to general eigenvalue problems where we need only moderate accuracy and can employ an eigenvalue routine which exploits the sparsity of the matrices generated by finite differencing.

Since the Nth eigenmode of the differential equation almost invariably oscillates too rapidly to be well approximated by N polynomials, the highest eigenfunctions of the equivalent N-dimensional matrix are purely numerical even when N is large enough so that the lowest modes and eigenvalues of the matrix are excellent approximations to those of the differential equation. Because Chebyshev series have the property of infinite order convergence, one can usually distinguish true modes from numerical modes by inspection. The coefficients c_n of the true modes will decrease very rapidly once the series has started to converge, i.e., for $n>M$ for some $M<N$, c_{n+1} will typically be one-half to one-fifth[9] the magnitude of c_n—this is the practical meaning of infinite order convergence. All the coefficients of the purely numerical modes in contrast will be of roughly the same order of magnitude. Thus, merely by printing out the coefficients of each matrix eigenfunction, one can usually tell at a glance which modes are numerical and which are good approximations to modes of the differential equation by keeping (2.6) in mind.

The safest and most reliable way of estimating the exact numerical value of the error is the time honored practice of repeating the calculation for a slightly larger or smaller N and comparing results. Boyd[8] discusses error estimates, pathological problems, and other aspects of pseudospectral method in more detail.

Given a column vector **F** containing the values of a function $f(x)$ at the collocation points (2.15), we can convert this to the coefficients of a truncated Chebyshev series (column vector **G**) and then into the coefficients of an ordinary polynomial (column vector H) on the standard interval of $[-1, 1]$ by two successive matrix multiplications.[10] The multiplication by the first matrix **R** merely evaluates the projection integrals (2.2) by Gaussian quadrature. The elements of the second matrix **Q** are the power series coefficients of each Chebyshev polynomial.

For our present purposes, the independent variable

is λ, the problem parameter, which for the anharmonic oscillator is a coupling constant that measures the strength of the x^4 term. The procedure begins by choosing an interval $[a,b]$ in λ and a degree N. We then solve the ordinary differential equation eigenvalue problem using the pseudospectral method in x for each of the N values of λ given by (2.15) and convert the eigenfunctions from the modified Chebyshev basis functions θ_n into ordinary Chebyshev polynomials by using (2.8) and (2.9). We can then multiply the N-dimensional column vector of values of each eigenvalue and each expansion coefficient of each eigenfunction by **R** and **Q** in turn to obtain Chebyshev and polynomial expressions for their dependence on λ. There is nothing complicated or subtle about this procedure: It is simply the rapid convergence of Chebyshev expansions that makes it practical as the anharmonic oscillator will illustrate.

Formally, to compute the expansions of $f(\lambda)$ on $[a,b]$ through N terms, we define the elements of the column vector F and the $N\times N$ square matrices **R** and **Q** as

$$F_i = f(x_i), \tag{2.18}$$

$$R_{ni} = 2T_{n-1}(y_i)/N, \quad n=1,2,\ldots,N, \tag{2.19}$$

$$Q_{11} = 0.5, \tag{2.20}$$

$$Q_{j1} = 0, \quad j=2,3,\ldots,N$$

$$Q_{jn} = \begin{cases} 0.5(n-1)(n-k-2)!\,2^{(j-1)}(-1)^k/(k!\,(j-1)!) \\ \text{if } j \text{ and } n \text{ are both odd or both even } \textit{and } j \le n, \\ 0 \quad \text{otherwise,} \end{cases}$$

$$n=2,3,\ldots,N, \; j=1,2,3,\ldots,N,$$

where

$$k \equiv (n-j)/2 \quad \text{(always an integer)} \tag{2.21}$$

and x_i is defined by (2.15) and y_i by (2.14). Then

$$f(\lambda) = \sum_{n=0}^{N-1}{}' G_{n+1} T_n\left(\frac{2}{(b-a)}\,[\lambda - 0.5(a+b)]\right) \tag{2.22}$$

$$= \sum_{n=0}^{N-1} H_{n+1}\left(\frac{2}{(b-a)}\,[\lambda - 0.5(a+b)]\right)^n, \tag{2.23}$$

$$G = \mathbf{R}F, \tag{2.24}$$

$$H = \mathbf{Q}G, \tag{2.25}$$

The prime on the sum in (2.22) denotes that the coefficient G_1 as computed by (2.24) should be taken with a factor of $\frac{1}{2}$ in evaluating the sum (as is automatically done by the recursive sum formula of Fox and Parker[4]). The reason is that the normalization coefficient [the C_n in (2.4)] for T_0 is twice that for the other modes, so (2.19) will compute a value for G_1 that is twice the actual coefficient of T_0. [Note that the coefficients of T_0 in (2.8) and (2.9) are the actual coefficients, and should be multiplied by two before using the Fox–Parker formula to evaluate the modified Chebyshev basis functions.] This is discussed further in the appendices.

As a very useful fringe benefit, we can repeatedly differentiate the truncated Chebyshev series and evaluate it at $\lambda = 0$ to obtain Taylor series representations of the function about $\lambda = 0$ (or any place else we want). If we denote the coefficients of a Chebyshev series by c_0, c_1, c_2, etc. and those of the first derivative by $c_0^{(1)}$, $c_1^{(1)}$, $c_2^{(1)}$, and so forth, then the two are related for a series truncated after N terms via (Fox Parker[4])

$$\begin{aligned} c_{N-2}^{(1)} &= 2(N-1)c_{N-1}, \\ c_{N-3}^{(1)} &= 2(N-2)c_{N-2}, \\ c_{i-1}^{(1)} &= 2(i)c_i + c_{i+1}^{(1)}, \quad i = N-3, N-4, \ldots, 1. \end{aligned} \tag{2.26}$$

We cannot blithely estimate the error by merely taking the absolute value of the first neglected term, however, because (2.26) shows that every other coefficient of the first derivative should contain $c_{N-1}^{(1)}$, which is the lowest neglected derivative coefficient. Thus, the error in the first derivative is roughly $(N/2)\,c_{N-1}^{(1)}$, not $c_{N-1}^{(1)}$ as we might expect. Differentiation is an antismoothing operation in general, so this rapid error propagation is not surprising. Because of it, we can only hope to calculate up to the second or third order Taylor coefficients with much accuracy, but for purposes of independently checking analytic perturbation theory calculations, this is all we need.

For reference in estimating the error in Taylor coefficient calculations, I will list explicit formulas for the coefficients of the first and second derivatives of a function in terms of those for the function itself:

$$\begin{aligned} c_n^{(1)} &= 2\sum_{j=0}^{\infty} (n+1+2j)c_{n+1+2j}, \\ c_n^{(2)} &= 4\sum_{m=0}^{\infty} (n+2+2m)c_{n+2+2m}\left(\sum_{k=0}^{m} (n+1+2m-2k)\right). \end{aligned} \tag{2.27}$$

3. APPLICATION TO THE ANHARMONIC OSCILLATOR

To illustrate these methods, I will apply them to parameterize the dependence of the eigenfunctions and eigenvalues of the anharmonic oscillator on the coupling constant λ. Formally, the problem is to find the solutions of the differential equation

$$\left[\frac{-d^2}{dx^2} + \frac{x^2}{4} + \frac{\lambda x^4}{4} - E(\lambda)\right] u(x) = 0, \tag{3.1}$$

which are bounded at $\pm\infty$ for all positive values of λ. For large values of λ, it is convenient to rescale the problem by setting $y = \lambda^{1/6}x$ to transform (3.1) into

$$\left[-\frac{d^2}{dy^2} + \frac{\lambda^{-2/3}}{4}\,y^2 + \frac{y^4}{4} - \lambda^{-1/3}E(\lambda)\right] u(y) = 0. \tag{3.2}$$

In the eigenfunction parameterizations I will give in terms of $\lambda^{-2/3}$, I will always use the rescaled spatial coordinate y.

Because of the infinite domain, the λx^4 term will always dominate the x^2 term for sufficiently large $|x|$ and will not be a small perturbation for *any* finite λ. On the other hand, if λ is small, then since the eigenfunctions decay exponentially for large $|x|$, this breakdown of the perturbative assumption (that λx^4 is small

TABLE I. Computed eigenvalues for the pure harmonic oscillator (symmetric modes only) sixty basis functions; domain [0, 15].

Mode No.	Eigenvalue	Mode No.	Eigenvalue
0	0.500000000827	2	2.49999999683
4	4.500000004721	6	6.50000000183
8	8.499999999427	10	10.50000000872
12	12.49999999725	14	14.50000000323
16	16.50000000462	18	18.49999998409
20	20.50000000358	22	22.50000001156
24	24.49999998710	26	26.50000001185
28	28.49999999720	30	30.49999999473
32	32.50000000341	34	34.50000001774
36	36.50000002619	38	38.50000019240
40	40.50000215247	42	42.50002129761
44	44.50016970882	46	46.50109803205
48	48.50570880576	50	50.52344871242

compared to x^2) will occur only where the eigenfunctions have negligible amplitude anyway. The mathematical consequences of these competing factors are, first, that $E(\lambda)$ *nonanalytic* at $\lambda=0$ and, second, that $E(\lambda)$ has an asymptotic power series about $\lambda=0$. As discussed theoretically in Appendix B and as demonstrated empirically by the numerical results that will be presented in a moment, the Chebyshev series for a smooth function with a nontrivial asymptotic series about one end point of the expansion interval converges almost as rapidly as that for a function which has no singularities of any kind near the expansion interval. The branch point at $\lambda=0$ which spoils the usefulness of the Rayleigh–Schrödinger series for $E(\lambda)$ for all but the smallest values of λ will cause no problems at all for the Chebyshev algorithm.

There are three different spectral methods for solving (3.1): Chebyshev polynomials on a large but finite interval (domain truncation), Chebyshev polynomials after a change of variable to transform $[-\infty,\infty]$ into $[-1,1]$ (mapping), and Hermite functions. All three work, but, for this paper, I will use domain truncation because it is the simplest and it shows most clearly that Chebyshev polynomials are a jack-of-all-trades, successful for a very wide range of problems without special tricks. Experiments with mapping were not too encouraging, largely because the singularity at the end point of the transformed interval is strong enough to significantly degrade convergence as discussed in Appendix B, but a different mapping might be more efficient. Hermite functions are quite successful for this problem since they are the exact eigenfunctions for $\lambda=0$, but there are a number of important algorithmic differences from Chebyshev polynomial methods which are discussed in Boyd.[12] The naive intuition that the last two approaches are *a priori* better than domain truncation because they employ infinite intervals is false. It is not possible in general to accurately represent the behavior of a transcendental function over an entire infinite interval with only a finite number of expansion functions: Though the *absolute* error of the approximation may be bounded by a small number, the *relative* error of a finite Hermite series for large x or of a Chebyshev series in the transformed variable y near $y=1$ will be enormous. Domain truncation merely faces up directly to this nonuniformity in x which the other two approaches hide.

The computational boundary b in the domain truncation approach should be chosen so that the eigenfunctions at $x=b$ have decayed by a factor roughly equal to the machine precision. Then, computer round off rather than the value of b is the primary source of error. Common sense, the known asymptotic solutions, and a little experimentation are usually adequate to choose a satisfactory b for this sort of problem: The exact choice is not crucial.

Since the operator of the differential equation commutes with the parity operator, the eigenfunctions are all of definite parity, symmetric or antisymmetric about $x=0$. To save computer time, we can compute these two classes of functions separately on the domain $[0,b]$ by imposing the boundary conditions $u(b)=0$ for all runs where b is a large positive number and set $u(0)=0$ to find the antisymmetric functions and $u'(0)=0$ to find the symmetric. In most of the results I present here, I set $b=10$ [for either (3.1) or (3.2)] and used forty Chebyshev basis functions, but I set $b=15$ and used sixty basis functions for the run whose results are presented in Table I.

The Chebyshev eigenfunction routine can compute only

TABLE II. The coefficients of the Chebyshev series and polynomial representations of the $n=0$ and $n=2$ energy levels on the interval $\lambda\in[0, 0.2]$.

	$n=0$		$n=2$	
Degree	Chebyshev	Polynomial	Chebyshev	Polynomial
0	1.105943311	0.55914632722	6.1641097992	3.1386243075
1	0.050564446	0.04887362253	0.5037752812	0.4694183996
2	−0.003964286	−0.00706976272	−0.0598379152	−0.0964126504
3	0.000607050	0.00195343470	0.0130682472	0.0356829649
4	−0.000122063	−0.00069716815	−0.0036119762	−0.0164273160
5	0.000028730	0.00028107050	0.0011378728	0.0071982175
6	−0.000007519	−0.00012653510	−0.0003903184	−0.0038097517
7	0.000002128	0.00008289358	0.0001422674	0.0066758966
8	−0.000000640	−0.00004453260	−0.0000542873	−0.0045442714
9	0.000000202	−0.00001141573	0.0000214634	−0.0044390767
10	−0.000000067	0.00000842004	−0.0000087012	0.0033076915
11	0.000000022	0.00002297234	0.0000035276	0.0036122576
12	−0.000000007	−0.00001416434	−0.0000012635	−0.0025875717

TABLE III. The coefficients of the Chebyshev series and polynomial representations of the $n=0$ and $n=2$ energy levels on the interval $\lambda^{-2/3} \in [0., 0.731]$ ($E_n = \lambda^{1/3} q_n$, where q_n is given by the appropriate polynomial in the variable $\lambda^{-2/3}$ whose coefficients are tabulated below).

	$n=0$		$n=2$	
Degree	Chebyshev	Polynomial	Chebyshev	Polynomial
0	1.7109068859	0.86181687593	10.766549418	5.3946719515
1	0.1806776547	0.17955804917	0.674890308	0.6742776808
2	−0.0063876680	−0.01258245023	−0.011392416	−0.0228226281
3	0.0003758331	0.00147189779	0.000202591	0.0008297053
4	−0.0000243347	−0.00018992200	0.000004909	0.0000354032
5	0.0000015866	0.00002488664	−0.000000977	−0.0000155690
6	−0.0000000999	−0.00000307919	0.000000083	0.0000023304
7	0.0000000059	0.00000003723	−0.000000005	0.0000008978
8	−0.0000000002	−0.00000020852	—	0.0000004538
9	—	0.00000032174	—	−0.0000012754
10	—	0.00000020296	—	−0.0000002982
11	—	−0.00000011226	—	0.0000004787
12	—	−0.00000007884	—	0.0000000728

a finite number of the lowest modes, but Hioe *et al.*[2] show that WKB calculations, which become more accurate as the mode number increases, agree with the exact results to many decimal places even for n as small as 10. Table I lists the computed symmetric eigenvalues for the pure harmonic oscillator; since the zeros and eigenvalues for the antisymmetric eigenfunctions interlace those of the symmetric, results for the antisymmetric are comparable. The exact eigenvalues are $(n+0.5)$, where n is a nonnegative integer. We see from the table that, with this choice of domain and basis size, we can compute the first forty eigenvalues to within a relative error of at most one part in 2×10^7. The situation for general λ is a little more complicated since the eigenfunctions are more closely confined to the origin as λ increases and thus the choice of optimum computational domain varies by about a third over the two intervals, $\lambda \in [0,0.2]$ and $\lambda^{-2/3} \in [0,0.731]$, that I used for the calculations I will present in the rest of this section; but, without resorting to out-of-core storage, it is possible to use as many as ninety basis functions on the CDC 7600 I employed for these computations. Thus, there is a large overlap between the Chebyshev and WKB methods By using the former for small n and the latter for large n, we can efficiently compute and explicitly represent the eigenvalues and eigenfunctions for all modes.

Since we have seen that the Chebyshev method is effective for a large number of modes, for the sake of simplicity only, I will confine the rest of my discussion to the ground state and the second excited state. Tables II and III present the coefficients of the Chebyshev and ordinary polynomial representations of the eigenvalues for these two modes on the intervals $\lambda \in [0, 0.2]$ and $\lambda \in [0.2, \infty]$ (in the variables λ and $\lambda^{-2/3}$, respectively). As discussed in the second paper by Hioe *et al.*,[2] the transition between small λ and large λ behavior occurs at smaller and smaller values of λ as the mode number increases. Thus, it is hardly surprising that the expansion for the second excited state converges more slowly than that for the ground state on $\lambda \in [0,0.2]$ but more rapidly on $\lambda \in [0.2,\infty]$. The polynomial form is probably easier to use (although the Chebyshev series can be summed in the same number of operations), but the Chebyshev form is useful because we can put a bound on the error of truncating the Chebyshev series after a given number of terms N (with $N<13$ here) by summing the absolute values of the neglected coefficients, as discussed earlier. For the ground state, for example, we find that two quadratic polynomials, one for $[0,0.2]$ and another for $[0.2,\infty]$, give the true eigenvalue to within a maximum relative error of 1 part in 700 and 1 part in 1500, respectively. The energy level formulas for Hioe *et al.*[3] are adequate for all practical purposes, and this present work would have little point if its only goal were to improve upon them. However, it is nonetheless a reflection of the power of the Chebyshev method that the largest error of either of these quadratics in the respective intervals is an order of magnitude smaller than those of the solutions of Hioe *et al.*'s cubic equation for $\lambda \in [0,0.2]$ and of their three term Taylor expansion in $\lambda^{-2/3}$ for $\lambda \in [0.2,\infty]$.

It has been shown that ground state eigenvalue of the anharmonic oscillator is a member of one of the few classes of functions for which it can be rigorously proven that the $[N,N]$ Padé approximants converge to the exact eigenvalue as N increases for all positive finite values of λ. Although Padé approximants are widely used in physics, it is impossible to prove convergence for most functions, so it is especially illuminating to compare Padé approximants with Chebyshev series for this problem. The $[N,N]$ Padé approximant is formally defined as that rational function whose numerator and denominator are polynomials of degree N and whose Taylor expansion agrees with that of the approximated function through the first $2N+1$ terms. By convention, the constant of the denominator polynomial is set equal to 1, so that the $[N,N]$ approximant has $2N+1$ undetermined coefficients which are computed by multiplying through by the denominator polynomial, formally matching powers of λ and solving the resulting set of linear equations. Like power series, Padé approximants are exact at the origin and become less and less accurate as λ increases, but empirically, even for functions with nonpolar singularities, the $[N,N]$ approximant usually has a much greater range of

TABLE IVa. A comparison of exact and approximate Taylor series about $\lambda = 0$ and $\lambda^{-2/3} = 0$ for the ground state.

		$\lambda = 0$	
Degree	Approximate	Exact	Relative error
0	0.50000000643	0.5	1.3×10^{-8}
1	0.749989	0.75	1.5×10^{-5}
2	−2.6218	−2.625	1.2×10^{-3}
3	20.416	20.8125	0.019
4	−214.	−241.289	0.113
		$\lambda^{-2/3} = 0$	
Degree	Approximate	Hioe *et al.*[2]	Relative error
0	0.66798626	0.66798626	0.
1	0.57468	0.57468	0.
2	−0.138	−0.141	0.021

usefulness than the truncated power series from which it was formed.

The [1,1] approximant about $\lambda = 0$ for the ground state eigenvalue has the same number of nontrivial coefficients as the quadratic Chebyshev series for [0,0.2]. By direct comparison, however, the Chebyshev series is more accurate in the uniform norm than the [1,1] approximant for $\lambda > 0.05$, being six times more accurate for $\lambda = 0.1$ and 25 times more accurate for $\lambda = 0.2$. The quadratic Chebyshev series is more accurate than the [2,2] approximant for $\lambda > 0.12$ and is five times more accurate at $\lambda = 0.2$. Thus from the perspective of obtaining an approximation which is accurate to within a specified tolerance over the entire interval [0,0.2], Chebyshev series are considerably more efficient than Padé approximants. Loeffel[13] *et al.* discuss the application of Padé approximants to the anharmonic oscillator in detail and give additional references.

The low order coefficients of the polynomial form of a high degree Chebyshev series will contain significant contributions from the higher Chebyshev terms because the high degree Chebyshev polynomials have constant, linear, quadratic, and higher terms with rather large coefficients as can be seen from their explicit power series (Abramowitz and Stegun[14]). Thus the polynomial form of a Chebyshev series should never be truncated; if we need less accuracy, we should truncate the Chebyshev coefficients and convert the result to polynomial form.

In the previous section, I made much of the fact that Chebyshev expansions converge on any interval on which the function is analytic. For $\lambda > 0.1$, the (asymptotic) Rayleigh–Schrödinger series about $\lambda = 0$ is essentially useless, but a quadratic polynomial on $\lambda \in [0,0.8]$ (not given in the tables) gave the ground state eigenvalue to within a relative error of one part in 80 over the whole interval. This is a very dramatic improvement over the Taylor series result, and illustrates the remark of the previous section that weak singularities at an end point will usually not seriously damage the convergence rate of the Chebyshev expansion.

Table IVa compares the Taylor expansions I computed from the Chebyshev series given in Table II and III with more accurate values from Hioe *et al.*[2] We see that the computed Taylor coefficients, although accuracy falls off rapidly with increasing order as explained in the previous section, are quite adequate for checking the first three or four terms of an analytically determined perturbation series.

To investigate the errors in differentiated Chebyshev series more carefully, I repeated the calculation of the ground state eigenvalue as a Chebyshev series of only seven terms and compared the first derivative of this with the first derivative of the 13-term expansion given in Table II. The results and the differences are given in Table IVb. Note the striking regularity in the absolute errors. As explained in the previous section, each even coefficient of the $N = 7$ expansion will be in error by a value roughly equal to the first neglected derivative coefficient and each odd coefficient will be in error by roughly the value of the second (first odd) neglected derivative coefficient. Curiously, the magnitude of the second neglected coefficient is somewhat smaller than the errors in the lower odd coefficients though of the right sign, but in general, Table IVb is a good illustration of the remarks about differentation errors made in the previous section. If we truncate the derivative of the 13-term expansion after six terms, we make only the truncation error given by the sum of the absolute values of the neglected derivative coefficients. If we truncate to seven terms *before* differentiation, then we will make errors in the computation of low-order derivative coefficients too. Round-off errors in the original, undifferentiated series will propagate through successive differentiations in much the same way as truncation errors, so I will not discuss this in detail.

The most important consequence of the swift convergence of Chebyshev series is that we can compactly parametrize the eigenfunctions themselves. For the anharmonic oscillator modes, we are cheifly interested in $|x| \leq 3$ since the amplitudes of the eigenfunctions are exponentially small for large absolute values of x. There are two ways of obtaining expansions valid for

TABLE IVb. A comparison of Chebyshev coefficients for the first derivative of the ground state eigenvalue for $\lambda \in [0.,0.2]$ as calculated using seven and 13 collocation points in λ and of the value of the first derivative at $\lambda = 0$.

First Taylor coefficient		
Exact	$N = 7$ approximation	Relative error
0.75	0.74798	0.0027

First derivative Chebyshev coefficients			
Degree	$N = 13$ ("exact")	$N = 7$	Absolute error
0	0.105092410	0.105056321	0.000036
1	−0.016935611	−0.016915633	−0.000020
2	0.003963518	0.003927436	0.000036
3	−0.001078467	−0.001058522	−0.000020
4	0.000321218	0.000285273	0.000036
5	−0.000101964	−0.000082551	−0.000020
6	0.00003392	—	—
7	−0.00001174	—	—

TABLE V. Coefficients of the double expansions of the ground and second excited states as Chebyshev series in x and λ for $\lambda \in [0, 0.2]$ (domain in x: $[0, 4.]$).

Degree in x	Degree in λ: 0	1	2
	$n=0$		
0	1.7404	−0.0387	0.0055
1	1.1254	0.0033	−0.0014
2	0.1866	0.0249	−0.0029
3	−0.1288	−0.0033	0.0016
4	−0.0625	−0.0066	0.0002
	$n=2$ state		
0	0.26733	−0.05100	0.00676
1	−0.90332	0.06823	−0.01029
2	−1.39567	0.00202	0.00542
3	−0.12335	−0.10855	0.01450
4	0.39571	0.01513	−0.00893
5	0.14547	0.03844	−0.00253
6	−0.03880	0.00033	0.00211
7	−0.02528	−0.00729	0.00002

TABLE VI. Coefficients of the double expansions of the ground and second excited states as Chebyshev series in x and $\lambda^{-2/3}$ for $\lambda^{-2/3} \in [0, 0731]$ (domain in x: $[0, 3.6]$).

Degree in x	Degree in $\lambda^{-2/3}$: 0	1	2
	$n=0$		
0	1.5795	−0.0358	0.0026
1	1.1236	−0.0086	0.0006
2	0.2979	0.0241	−0.0020
3	−0.1212	0.0146	−0.0006
4	−0.1000	−0.0061	0.0008
	$n=2$		
0	0.0716	0.0121	−0.0017
1	−0.6510	0.0337	−0.0023
2	−1.3002	0.0192	−0.0002
3	−0.5578	−0.0383	0.0024
4	0.3224	−0.0309	0.0003
5	0.3328	0.0141	−0.0015
6	0.0204	0.0188	−0.0003
7	−0.0656	−0.0001	0.0006

this small interval: The first is to calculate the eigenfunctions on a large domain and then re-expand them on a small domain and the second is to simply compute the modes *ab initio* with boundaries at small absolute values of x. The first is a little more accurate at the expense of a little extra programming, so to illustrate the power of the Chebyshev methods, I have used the second, employing the x domains $[0,3.6]$ and $[0,4.0]$, respectively, for the intervals $[0,0.2]$ and $[0,2,\infty]$ in λ.

The coefficients of the double Chebyshev expansions, to second order in λ or $\lambda^{-2/3}$ and to fourth order in x for the ground state and to seventh order in x for the excited state, are given in Tables V and VI. Table VII compares the values of these truncated expansions with the exact eigenfunctions for $x \leq 3$ and particular values of λ and $\lambda^{-2/3}$. The largest absolute errors are only 1% of the peak amplitude of the wavefunction. Note that the errors alternate in sign and that the error at several points is roughly equal in absolute value to the largest error at any point. This is a consequence of the strong resemblance of truncated Chebyshev approximations to the best approximating polynomial in the uniform norm as was mentioned earlier. Because of this highly uniform nature of Chebyshev expansions, the errors for values of λ and $\lambda^{-2/3}$ other than those listed in the table where roughly of the same magnitude.

We see that by using two different expressions of fifteen terms, one for small coupling constant, the other for large, we can represent the ground state wavefunction to moderate accuracy for all positive values of λ and all values of x where the mode has significant amplitude. In a numerical sense, we can consider the anharmonic oscillator problem as completely solved and its eigenfunctions as known.

4. DISCUSSION AND SUMMARY

The anharmonic oscillator is a simple problem which, because of its simplicity, can be attacked using any number of special tricks, but the real significance of my results is twofold. First, the two-part Chebyshev algorithm should refute the notion that if one wants the answer in the form $E(\lambda) = E_0 + E_1\lambda + E_2\lambda^2 \cdots$, one must inevitably use some form of perturbation theory. Second, the double algorithm is completely general. In particular, Chebyshev methods can be used to solve boundary value problems, integral equations, and more exotic classes of equations as discussed in Fox and Parker[4]; I have limited the discussion to eigenvalue problems only for clarity. I have successfully used the programs that generated the results of the previous section on a very different problem without modification: the generalized Laplace's tidal equation, which (after separation of variables) describes the north–south structure of atmospheric planetary waves in the

TABLE VII. A comparison of exact and approximate values for the $n=0$ and $n=2$ eigenfunctions for $x \in [0, 3.]$ and $\lambda^{-2/3} = 0.3655$, $\lambda = 0.1$.

x	$n=0$ Approx.	Exact	Error	$n=2$ Approx.	Exact	Error
		$\lambda^{-2/3} = 0.365$				
0	0.995	1.000	0.005	−0.929	−0.932	−0.003
0.33	0.957	0.953	−0.004	−0.669	−0.667	0.002
0.67	0.816	0.821	0.005	−0.018	−0.024	−0.006
1.00	0.624	0.631	0.007	0.633	0.635	0.002
1.33	0.426	0.423	−0.003	0.971	0.980	0.009
1.67	0.251	0.242	−0.009	0.921	0.917	−0.004
2.00	0.118	0.114	−0.004	0.621	0.611	−0.010
2.33	0.037	0.043	0.006	0.294	0.299	0.005
2.67	0.004	0.013	0.009	0.097	0.106	0.009
3.00				0.036	0.027	−0.009
		$\lambda = 0.1$				
0	0.995	1.00	0.005	−0.909	−0.907	0.002
0.33	0.974	0.969	−0.005	−0.754	−0.754	0.000
0.67	0.882	0.882	0.000	−0.348	−0.347	0.001
1.00	0.748	0.752	0.004	0.177	0.173	−0.004
1.33	0.595	0.597	0.002	0.645	0.642	−0.003
1.67	0.443	0.442	−0.001	0.932	0.932	0.000
2.00	0.305	0.302	−0.003	0.995	0.996	0.001
2.33	0.191	0.189	−0.002	0.871	0.872	0.001
2.67	0.107	0.107	0.000	0.643	0.647	0.004
3.00	0.052	0.055	0.003	0.403	0.413	0.010

presence of a mean (independent of longitude) wind that is varying with latitude. The geometry is spherical; there are true singularities at the poles and "apparent" singularities near the equator; the problem is different for each wind profile; and because of the geometric, Coriolis force, and wind-dependent terms, the coefficients of the differential equation are too complicated to write down explicitly and must be computed in stages. Nonetheless, both parts of the Chebyshev algorithm work well, and I will soon publish results for this class of problems in another place.

It is really for such more difficult problems that general purpose algorithms such as I have described are intended. An almost embarrassing number of special techniques have been applied to the anharmonic oscillator energy levels, and as usual some of these are computationally cheaper or more accurate in limited domains than results of the most general methods—Hioe *et al.*'s[2] Taylor expansions in $\lambda^{-2/3}$, for example, for $\lambda \gg 1$. Algorithms can never substitute for thinking. Still, the Chebyshev methods have made it possible to give new results for this venerable war horse: compact, explicit polynomial representations for the eigenfunctions. This speaks for itself.

ACKNOWLEDGMENTS

I wish to thank Dr. Cecil E. Leith and an anonymous referee for reading this work and suggesting useful improvements. This research was performed while I was a Postdoctoral Fellow of the Advanced Study Program of the National Center for Atmospheric Research, which is sponsored by the National Science Foundation. Computations were performed on the CDC 7600 at this institution.

APPENDIX A: RECURSIVE CHEBYSHEV FORMULAS

To evaluate the Chebyshev polynomials and their first and second derivatives efficiently, we can use the formulas (prime denotes differentiation)

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \quad n = 0, 1, \cdots, \tag{A1}$$

$$T_n'(x) = 2x[n/(n-1)]T_{n-1}'(x) - [n/(n-2)]T_{n-2}'(x),$$

$$n = 3, 4, \cdots, \tag{A2}$$

$$T_n''(x) = 2x[n/(n-2)]T_{n-1}''(x) - [n^2/(n-2)^2]T_{n-2}''(x),$$

$$n = 3, 4, \cdots \tag{A3}$$

To make (A1) valid for all nonnegative integral values of n, we define

$$T_{-1}(x) \equiv x,$$

$$T_{-2}(x) \equiv 2x^2 - 1 \tag{A4}$$

for computational convenience. The true Chebyshev polynomials are defined only for nonnegative integral values of n. To supplement (A2) and (A3) which are valid only $n \geq 3$, we use the starting values

$$T_0'(x) = 0, \quad T_1'(x) = 1, \quad T_2'(x) = 4x, \tag{A5}$$

$$T_0''(x) = 0, \quad T_1''(x) = 0, \quad T_2''(x) = 4. \tag{A6}$$

To compute the sum of a function $f(x)$ where $x \in [a, b]$ as a truncated series of Chebyshev polynomials of the stretched variable y where $y \in [-1, 1]$ (the standard interval), i.e.,

$$f(x) = \sum_{n=0}^{N-1}{}' c_n T_n(y), \tag{A7}$$

we initialize by setting

$$y = 2(x - (a+b)/2)/(b-a)$$
$$b_1 = 0, \quad b_2 = 0, \tag{A8}$$

and then cycle through the following loop N times:

$$b_0 = 2y b_1 - b_2 + c_{N-i}, \quad i = 1, \ldots, N,$$
$$b_3 = b_2, \quad b_2 = b_1, \quad b_1 = b_0. \tag{A9}$$

At the end of the loop

$$f(x) = 0.5(b_1 - b_3). \tag{A10}$$

The prime on the summation in (A7) denotes that the lowest term in the sum is to be taken with a factor of $\frac{1}{2}$, i.e., $c_0 T_0/2$ is to be added to the sum of the other terms, not $c_0 T_0$, and the algorithm of (A8)–(A10) automatically does this. As explained just after (2.25), this pesky factor of two stems from the normalization factors C_n in (2.2), which are

$$c_0 = \pi, \quad c_n = \pi/2. \tag{A11}$$

One can either divide the first row of the matrix defined in (2.19) by two, or multiply the computed c_0 by 0.5 in summing the Chebyshev series. Following Fox and Parker,[4] I have chosen the latter. A similar problem with the constant term arises in ordinary Fourier series, which is hardly surprising since (2.7) shows that the Chebyshev polynomials are merely cosine functions of a transformed variable.

APPENDIX B: THE EMPIRICAL CONVERGENCE RATE OF A CHEBYSHEV SERIES

If a function is infinitely differentiable but nonanalytic at a point, it still has a formal power series expansion about that point, and it can be shown through the Taylor series remainder formula that this series is asymptotic to the function. It can further be shown that the *Chebyshev* expansion of the function on an interval which includes this point as one end point still has the property of *infinite-order convergence*. The proof follows from making the change of variable $x = \cos\theta$, using the relationship between the trigonometric functions and Chebyshev polynomials, and integrating the explicit expression for the series coefficients (2.2) by parts k times to show that the coefficients are $O(n^{-k})$ as n goes to infinity, provided the function is k times differentiable. If the function is infinitely differentiable everywhere on the interval of expansion including the end points, then the Chebyshev series will show infinite order convergence even if it is not everywhere analytic.

It is still important, however, to gain some feeling, even if only on an empirical basis, for the practical convergence rates of Chebyshev series for functions which are singular at one end point although infinitely differentiable there. The expansion with the coefficients $a_n = (0.99)^n$ technically has the property of infinite order convergence, but one would need to sum several thou-

TABLE VIII. The ratios $r_n = |a_n/a_{n+1}|$ for the $n=0$ eigenvalue for $\lambda \in [0, 0.2]$.

n	r_n
0	21.86
1	12.75
2	6.53
3	4.97
4	4.25
5	3.82
6	3.53
7	3.33
8	3.17
9	3.01

sand terms to compute the function accurately, thus showing that this theoretical property is no guarantee of a series' usefulness. Luke,[15] however, tabulates the expansion coefficients for a number of functions which are weakly singular at one end point, and finds rapid convergence, just as is true here for the anharmonic oscillator as shown in my tables. For a couple of series, he demonstrates that the coefficients are proportional to $\exp(-pn^q)$ for a large n where q is $\frac{1}{2}$ or $\frac{2}{3}$, thus explicitly illustrating the infinite order convergence.

The most useful measure of the rate of convergence of a Chebyshev series is $r_n = |a_n/a_{n+1}|$. For most functions, r_n is usually between two and five once n is large enough to give moderate (5%) accuracy. For functions whose only singularities are simple poles, it can be shown rigorously that r_n is asymptotically *constant*. For entire functions—exponentials, sines, and Bessel functions—r_n asymptotically *increases*. In Luke's asymptotic formulas, however, r_n will *decrease* to one from above as n tends to infinity if q is less than one, as it is for his examples. Table VIII shows that the same behavior is true of the Chebyshev series for the ground state eigenvalue of the anharmonic oscillator on an interval with one end point at $\lambda = 0$.

It is straightforward to show that r_n must decrease to one as $n \to \infty$ if the function is singular at one or both end points. It is easy to prove that a Chebyshev series will converge inside the largest ellipse with foci at the end points of the expansion interval within which the function is analytic. (As demonstrated in Geddes,[16] the same conformal mapping that transforms an ellipse with foci at ± 1 into a circle about the origin also maps $T_n(z)$ into w^n. Thus, the convergence of the Chebyshev series for $f(z)$ is equivalent to that of the Taylor series in w for $g(w) = f(z)$, from which the theorem follows trivially.) For functions which are singular at one or both end points, the ellipse of convergence collapses into the expansion interval itself: The series diverges at all other points of the complex plane. Boyd[17] shows that Chebyshev polynomials satisfy the tight bound

$$|T_n(x+iy)| \leq \cosh n\mu_0 \tag{B1}$$

for all points inside and on the ellipse $\mu = \mu_0$, where μ is an elliptical coordinate in the complex plane. (Note that the ellipse for $\mu_0 = 0$ is the expansion interval itself.) Thus, the ratio

$$|a_n T_n(x+iy)| / |a_{n+1} T_{n+1}(x+iy)|$$

is bounded from below by $r_n \exp(-\mu_0)$. If r_n does *not* asymptotically tend to 1, then for sufficiently small μ_0, the Chebyshev series must (by the ratio test) converge uniformly and absolutely inside and on the ellipse $\mu = \mu_0$, which is a contradiction. Divergence of the series at all points off the expansion interval is not incompatible with the property of infinite order convergence, however. If the coefficients are asymptotically proportional to $\exp(-pn^q)$, for example, one clearly has infinite order convergence on the expansion interval, but for $\mu_0 \neq 0$ and for $q < 1$, the $\exp(n\mu_0)$ growth of the Chebyshev polynomials will eventually overwhelm the slower exponential decrease of the absolute values of the coefficients, implying that the series diverges as it must.

On a practical level, however, the importance of the end point singularity depends on its severity. The function $\exp[-0.25 \log^2(1-y)]$, which is the ground state eigenfunction of the harmonic oscillator after the change of variable $y = 1 - \exp(-x)$ is made in (3.1), has only the trivial asymptotic series about $y = 1$, i.e., $f(y) = 0$, since all the derivatives of the function vanish there. Nonetheless, the integrals defining the Chebyshev coefficients can be integrated by parts an arbitrary number of times, and Chebyshev series with $N = 3$, 24, and 49 give maximum absolute errors of 0.033, 2.2×10^{-4}, and 2.3×10^{-6}, respectively. In noteworthy contrast to the usual situation where the errors of a truncated Chebyshev series are distributed more or less uniformly over the interval, for this function, they are concentrated near the singularity. Away from the neighborhood of $y = 1$, for example, the maximum errors are less than those at $y = 1$ by a factor of four for $N = 3$ and by a factor of roughly one hundred for $N = 24$ and $N = 49$. Clearly Chebyshev series give useful and highly accurate approximations to this function, especially if one is willing to tolerate higher error near the singularity, but convergence is slow in comparison with that for a monotonic analytic function. It is because of this slow convergence that domain truncation, instead of a change of variable to map $[0, \infty]$ into $[0, 1]$, was used to compute the eigenfunctions of the anharmonic oscillator.

When the singularity is weaker so that the asymptotic series about the singularity is nontrivial and has some small but finite range of usefulness, then the asymptotic behavior of r_n is probably significant only if one is interested in very high accuracy—more than ten decimal places—for most simple functions. For

$$e^x x^{1/2} \int_x^\infty K_0(t)\,dt,$$

r_{25} is still as large as 3.16 as compared to $r_{10} = 4.83$; $n = 10$ gives nine decimal place accuracy and $n = 25$ gives 18 decimal place accuracy for this function. Similarly, $r_9 = 3.01$ for the ground state eigenvalue of the anharmonic oscillator; the first eleven terms of this series give a relative error of only 1 part in 10 million. Thus, even though r_n is tending to one for large n for both these examples, r_n is still within the empirical range of "two to five" that I quoted above for the largest values of n that are likely to be of practical interest. Empirically, then (as well as theoretically), the Chebyshev expansions of functions which are singular

but infinitely differentiable at one end point and which have nontrivial asymptotic expansions about that point are likely to be as useful and almost as rapidly converging as those of smooth, well-behaved functions with no singularities of any kind close to the expansion interval.

Notes added in proof: (i) Grosch and Orszag[18] have thoroughly investigated mappings for six examples including the harmonic oscillator ($\lambda = 0$). In general, they find that exponential mappings similar to that discussed in Appendix B are poorer than simple domain truncation for the same reason given here: The endpoint singularity in the transformed variable is strong enough to drastically reduce the convergence rate of the Chebyshev series. However, they also show that algebraic mappings of the form $y = x/(x + L)$ where L is a constant are usually somewhat more efficient than the domain truncation adopted for simplicity here.

(ii) Sheorey[19] in a paper published after the original submission of this manuscript has independently applied the same idea to the Coulomb wavefunctions. Although his emphasis is strictly numerical in contrast to the philosophy which I have stressed here, which is that the double Chebyshev algorithm is an analytic alternative to perturbation theory, his work should be noted.

(iii) The traditional philosophy of table-making is to calculate the solution at from twenty-five to one hundred *evenly* spaced values of λ and then print the results. For representing observational or experimental data, this strategy is sound, but for analytic functions, it is probably highly inefficient. It is unlikely that any table-maker striving for eight decimal place accuracy would construct a table with as few entries as Table II, but from the thirteen Chebyshev coefficients given there for $E_0(\lambda)$, one can in fact compute the function to at least eight figures over the entire interval. By using the double Chebyshev method, one can easily make a table with a *hundred* entries by doing the primary calculation (finding the eigenvalues via the QR algorithm) for only a *dozen unevenly* spaced values of λ.

[1]W. Kauzmann, *Quantum Chemistry: An Introduction* (Academic, New York, 1957).
[2]F. T. Hioe and E. W. Montroll, J. Math. Phys. **15**, 1945 (1975); F. T. Hioe, D. MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).
[3]D. M. Young and R. T. Gregory, *A Survey of Numerical Mathematics* (Addison-Wesley, Reading, Mass., 1972), Vol. I.
[4]L. Fox and I. B. Parker, *Chebyshev Polynomials in Numerical Analysis* (Oxford U. P., Oxford, 1968).
[5]J. Todd, *Introduction to the Constructive Theory of Functions* (Academic, New York, 1963).
[6]M. J. D. Powell, Comp. J. **9**, 404 (1967).
[7]S. Orszag, J. Fluid Mech. **50**, 689 (1971).
[8]J. P. Boyd, submitted to J. Comp. Phys., 1977.
[9]This range is representative based on empirical experience. It is not rigorous and the values given are somewhat arbitrary.
[10]One can replace the first matrix multiplication by using the FFT algorithm, which is faster, as described in Geddes,[11] but since this step is only a small part of the over-all computation, use of the slower but simpler matrix multiplication will not significantly reduce efficiency.
[11]K. O. Geddes, SIAM J. Numer. Anal., to appear (1977).
[12]John P. Boyd, J. Comp. Phys. to appear (1978).
[13]Loeffel *et al.*, Phys. Lett. B **30**, 656 (1969).
[14]M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions* (Dover, New York, 1965).
[15]Yudell Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. II.
[16]K. O. Geddes, *Theory of Approximation with Applications*, edited by A. G. Law and B. N. Sahney (Academic, New York, 1976).
[17]J. P. Boyd, submitted to J. Inst. Math. Appl. 1976.
[18]C. E. Grosch and S. A. Orszag, J. Comput. Phys. **25**, 273 (1977).
[19]V. B. Sheorey, Comput. Phys. Commun. **12**, 125 (1976).

The homotopy groups of condensed matter physics

N. D. Mermin

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853
(Received 28 December 1977)

The classes of distinct point and line defects in ordered media and the laws governing their combination, are determined by the fundamental group of the order parameter space π_1, the second homotopy group π_2, and the group of automorphisms induced in π_2 by the action of π_1. A simple algorithm is given for computing all of these algebraic structures for a quite general family of ordered forms of condensed matter.

I. INTRODUCTION

Several papers have recently appeared applying homotopy theory to the study of defects in the ordered media of condensed matter physics.[1-10] In such treatments the relevant properties of the media are to be found in the topological structure of the space of allowed values of its order parameter. Of particular importance are those topological features described by the structure of the first and second homotopy groups, π_1 and π_2. The categories of distinct line defects are determined by the algebraic structure of π_1 (known also as the fundamental group of the order parameter space). The categories of distinct point defects are determined by the algebraic structure of π_2, together with a topologically determined group of automorphisms of π_2 associated with the transport of point defects around closed paths encircling line defects, and characterized mathematically as being brought about "by the action of π_1 on π_2."

The primary concern of most of these papers has been how, given the structure of π_1 and π_2, one can characterize the classes of line and point defects, their combination laws, or what comes about when attempts are made to pass such defects through or around one another. Except in a few papers[9,10] the actual computation of these groups and their interrelations has not been of central concern; the groups have been determined in the following ways:

(a) Reference has been made to results in the mathematical literature;

(b) Ingenious arguments have been used to construct the groups from first principles, by explicitly considering all relevant deformations of curves or surfaces in the order parameter space;

(c) Powerful general theorems (summarized in the "exact homotopy sequence of homomorphisms") have been exploited to reduce the structure of the homotopy groups to a small number of possibilities. All but one of them is then eliminated by arguments essentially of type (b), applied to a much simpler set of cases than would have been required were the exact sequence not invoked.

I wish to point out that at least in the matter of computing the homotopy groups, one does not need a mathematics library, ingenuity, or facility in the manipulation of exact sequences. There is a simple algorithm which reduces to an elementary algebraic exercise the computation of the first and second homotopy groups, as well as the computation of the action of π_1 on π_2, for a very large class of ordered media. The class includes all those discussed to date in the literature[11] and many yet to be discussed. The presentation of this algorithm is organized as follows:

In Sec. II I characterize those ordered media to which the algorithm can be applied. The algorithm itself is presented in the form of three short theorems. It is described in enough detail to indicate in elementary terms its intuitive content; most of the comparatively unilluminating formal proof is deferred to Sec. IV.

In Sec. III I apply the algorithm in several cases, to provide some concrete illustrations of the rather abstract procedure specified in II. Readers more interested in how to use the algorithm than in why it works, might be advised to skip directly to Sec. III after reading the statement of the three theorems.

In Sec. IV I offer a brief proof of the algorithm by a single application of the exact homotopy sequence to all of the cases characterized in Sec. II. Readers with only a little less knowledge of topology than I have may fail to find the argument illuminating, and they are encouraged to construct their own proofs along the more elementary lines hinted at in Sec. II. I have chosen to present here the formal proof rather than pursuing the more intuitive line of argument, because the former is far more concise and emphasizes that the computational algorithm I am describing is, in fact, only a very modest extension of the approach [(c) above] already emphasized by Volovik and Mineyev.[7]

II. COMPUTING π_1, π_2, AND THE ACTION OF π_1 ON π_2

The media for which the algorithm is designed are those in which the order parameter space can be characterized as a space of cosets G/H of a Lie group G, with respect to a closed subgroup H. This description of the order parameter underlies the Landau–Lifshitz theory of phase transitions (and survives the failure of the more quantitative aspects of that theory). It can also be found in the general treatments of Refs. 3 and 7. To describe an order parameter space in this way one requires only the following:

There must be a group G of transformations among the possible values of the order parameter, which is large enough to contain operations taking any possible value into any other. Given such a group, one can single out a particular reference value ϕ_0 of the order param-

 0022-2488/78/1906-1457$1.00

eter ϕ, and specify any other value ϕ_1 by giving a transformation g in G which takes ϕ_0 into ϕ_1. In general there will be many elements of g satisfying $g\phi_0 = \phi_1$, but there is a simple way to deal with this redundancy. One notes that the set H of all transformations in G that leave ϕ_0 invariant, is a subgroup (known as the isotropy subgroup or "fixer" of the order parameter). It is then an elementary algebraic exercise to establish that the set of all transformations in G that take ϕ_0 into any given ϕ_1 is simply a left coset of H in G. One thus establishes a correspondence between distinct values of the order parameter, and the left cosets of the isotropy subgroup H in G.

The abstractness of this presentation should be relieved by an examination of the examples in Sec. III. (One might also note that in the Landau–Lifshitz language, G is the symmetry group of the disordered state and H is that subgroup of G that survives the symmetry breaking below the transition temperature.)

We call the space of left cosets G/H, and identify it with the order parameter space. The central observation which reduces the computation of the homotopy groups to a simple algebraic exercise, is that the group G can always be chosen so that $\pi_n(G)=0$ for $n=0,1$, and 2. In actual practice the choice of a connected $[\pi_0(G)=0]$ simply connected $[\pi_1(G)=0]$ group G is quite routine, and $\pi_2(G)$ vanishes automatically. Readers who are content with this are urged to skip from here to the statement of the algorithm (Theorems 1–3 below) and thence to the illustrative applications in Sec. III. The three paragraphs that follow indicate why these restrictions are easy to satisfy.

$\pi_0(G)=0$: This is homotopy-theoretic shorthand for the assertion that the continuous group G is arcwise connected. This can always be arranged for the study of point and line defects, for in the absence of planar defects that disconnect macroscopic portions of physical space, any two points of space can be joined by a path that avoids all order parameter singularities. The values of the order parameter at such a pair of points can then be related by a transformation of G that is in the connected component of the identity. Thus if one did happen to choose a disconnected group G [for example by taking G to be the group O(3) of proper *and* improper rotations, rather than the connected proper rotation group SO(3)], one could build a new description based only on operations in the subgroup of G connected to the identity, provided only that the region of physical space free of defects was connected.

$\pi_1(G)=0$: Such groups are said to be simply connected, and are characterized by the fact that any continuous one-parameter family of group elements $g(\theta)$ with $g(0) = g(2\pi)$(i.e., any continuous image of a circle in the group G) can be continuously deformed to a single point in G. If the original choice of G is not a simply connected group, one can always find a larger simply connected group, of which G is the homomorphic image, and work with this group instead. The existence of such a group (known as the "universal covering group") is guaranteed by a fundamental topological theorem.[12]

In applications to the media of condensed matter physics the selection of a simply connected group G presents no problems. The three dimensional proper rotation group SO(3) is not simply connected, but it is the homomorphic image of the simply connected group SU(2) of unitary unimodular 2×2 matrices; the two dimensional proper rotation group SO(2) is not simply connected, but it is the homomorphic image of the simply connected one-dimensional translation group T(1); the translation group T(n) is simply connected in any number of dimensions. The groups G encountered in applications in condensed matter physics are invariably built out of direct products (or, in the case of the Euclidean group, semidirect products) of the above groups. Since the product of two simply connected spaces is also simply connected, a simply connected G can always be arranged. In practice one merely describes rotations by operations from SU(2) or T(1), rather than SO(3) or SO(2). The additional redundancy in description that this introduces is no problem, since the formation of the coset space G/H removes whatever redundancy is present. The simplicity of the procedure should be revealed by the examples in Sec. III.

$\pi_2(G)=0$: This is the assertion that any continuous image of the surface S_2 of a 3-sphere in G can be continuously deformed to a single point. This is not difficult to prove for any of the groups described in the preceding paragraph, but in fact it is never a worry, due to a theorem of Cartan that $\pi_2(G)$ vanishes for any Lie group.[13]

We now state the algorithm in the form of three theorems. In all cases G is to be taken to be connected and simply connected. [The vanishing of $\pi_2(G)$, which is used in the proof of theorem 2, is automatic.]

Theorem 1: If G is connected and simply connected then $\pi_1(G/H)$ is isomorphic to the quotient group H/H_0, where H_0 is the connected component of the identity in H.

Theorem 2: If G is connected and simply connected, then $\pi_2(G/H)$ is isomorphic to $\pi_1(H_0)$.

Theorem 3: If G is connected and simply connected then the action of $\pi_1(G/H)$ on $\pi_2(G/H)$ is given by the action on $\pi_1(H_0)$ of the inner automorphisms of H. [This concise but somewhat cryptic statement can be expanded as follows: Let α be a homotopy class in $\pi_1(G/H)$ and let β be a homotopy class in $\pi_2(G/H)$. Let h be any element of H such that hH_0 is the coset of H_0 in H corresponding to α under the isomorphism of Theorem 1, and let g_t be any loop in H_0 belonging to the homotopy class corresponding to β under the isomorphism of Theorem 2. Then the action of α on β is to transform β into the homotopy class β' which corresponds, under the isomorphism of Theorem 2, to the homotopy class of loops in H_0 of which $h^{-1}g_t h$ is a representative. The resulting automorphism of $\pi_2(G/H)$ is independent of the choice of coset representative h or representative loop g_t.]

I expand briefly on the meaning and intuitive content of these results and then turn (in Sec. III) to some applications which will probably clarify their meaning more efficiently than any general remarks. Theorem 3 is proved in the discussion at the end of this section,

but the proof of Theorems 1 and 2 is deferred to Sec. IV. For applications of these results to the study of defects, the informal discussion in this section is likely to be of more interest than the rigorous proofs of the isomorphisms in Sec. IV. It is only in the informal discussion that the explicit construction of the isomorphic correspondences is specified. Without an understanding of that correspondence it can be difficult to sort out which classes of defects correspond to which elements of the abstract homotopy groups.

Comments on Theorem 1: The statement of Theorem 1 presupposes the elementary result that if H is any continuous group and H_0 the connected component of the identity in H, then H_0 is a normal subgroup of H, so that the coset space H/H_0 is, in fact, a group.[14] From the topological point of view the cosets of H_0 in H are the various disjoint connected components of H. The elements of the group H/H_0 are best thought of as these internally connected pieces of H, which are given a group structure by virtue of H_0 being a normal subgroup. (The product of two pieces is simply the piece to which the product of any pair of their respective members belongs.) The isomorphic mapping between $\pi_1(G/H)$ and the elements of H/H_0 can be characterized as follows:

Given any connected component H_1 of H, trace out a continuous path $g(t)$ $(0 \leq t \leq 1)$ that connects any point in H_1 to the identity e (which lies in H_0). The continuous family of cosets given by $K(t) = g(t)H$ is a closed path (i.e., a loop) in the coset space G/H [since $g(0)H$ and $g(1)H$ are both the coset H itself]. The homotopy class of loops to which the loop $K(t)$ belongs is the element in $\pi_1(G/H)$ associated with the element of H/H_0 given by H_1. A line singularity characterized by an element of $\pi_1(G/H)$ corresponding to the component H_1 of H can be represented as one in which the variation in the order parameter about an encircling path is given by $g_t\phi_0$, $0 \leq t \leq 1$.

A rigorous proof of Theorem 1 must establish that the structure of this correspondence is independent of the particular choice of paths from the components H_i of H to the identity, and that it meets the requisite features of being one-to-one and taking the multiplication table of the one group into that of the other. The proofs of all these points are straightforward consequences of the underlying definitions, and can be formulated along lines requiring no advanced topological technology. Once the connection between paths linking the H_i to H_0 in G, and closed loops at the coset H in G/H is firmly perceived, the proof of the theorem is virtually at hand.

Comments on Theorem 2: The structure of the isomorphic correspondence between $\pi_2(G/H)$ and $\pi_1(H_0)$ is the following:

The second homotopy group of the coset space G/H consists of homotopy classes of maps of the surface S_2 of a sphere into G/H, which take one point of the sphere (here chosen to be the south pole) into the coset H itself. If we introduce spherical coordinates then a representative map of a sphere into G/H is a two parameter family of cosets $K(\theta,\phi)$, with $K(\pi,\phi)$ identically equal to H. To prove Theorem 2 one must first establish that such a family of cosets can always be represented by a two-parameter family $g(\theta,\phi)$ of elements of G itself:

$$K(\theta,\phi) = g(\theta,\phi)H, \tag{2.1}$$

where $g(\pi,\phi)$ traces out a closed curve in H_0, as ϕ varies from 0 to 2π. [It is evident that such a map of a disk into G taking the circumference into H_0 represents, via (2.1), the image of a sphere in the coset space G/H; the nontrivial part is that any sphere in coset space can be so represented.]

Having established the representation (2.1) for any sphere at H in coset space G/H, one must then prove: (a) That the correspondence (2.1) is one-to-one between homotopy classes of maps of S_2 into G/H taking the south pole into H, and homotopy classes of maps of the disk into G taking the circumference into H_0; (b) That the homotopy classes of maps of the disk into G are entirely characterized by the action of the map on the circumference of the disk—i.e., by the elements of $\pi_1(H_0)$, which *are* the homotopy classes of loops in H_0; and (c) that the resulting correspondence between maps of spheres into G/H and circles into H_0 preserves the multiplication table between the homotopy classes in $\pi_2(G/H)$ and the homotopy classes in $\pi_1(H_0)$.

Proof of Theorem 3: Given the isomorphic correspondences described above, Theorem 3 follows from a careful statement of precisely how $\pi_1(G/H)$ acts on $\pi_2(G/H)$. Let $K(\theta,\phi)$ be a sphere in coset space represented by the disk $g(\theta,\phi)$ in G [as in Eq. (2.1)]. Let $L(\beta)$, $0 \leq \beta \leq 2\pi$, be a loop at H in coset space—i.e., a map of a circle into G/H [with $L(0) = L(2\pi) = H$] representing a homotopy class of $\pi_1(G/H)$. Under the correspondence given by Theorem 1 we can represent $L(\beta)$ by a path $a(\beta)$ in the group G connecting an element in one of the components of H to e:

$$L(\beta) = a(\beta)H, \quad a(0) = h_1 \in H_1, \quad a(2\pi) = e. \tag{2.2}$$

The loop L acts on the sphere K as pictured in Fig. 1, to produce a new sphere LK. Formally, LK is defined by

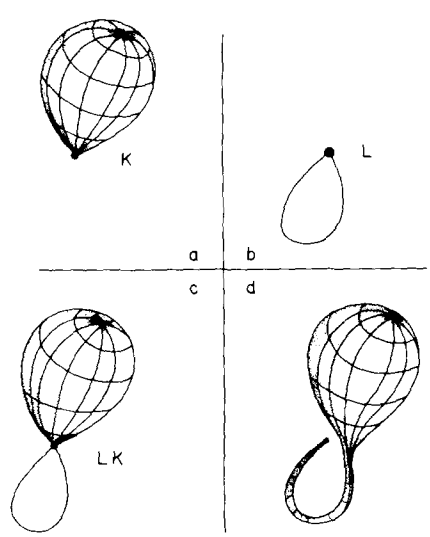


FIG. 1. Various sets in the coset space G/H. The heavy dot represents the coset H. (a) The image K of a sphere at H. (b) The image L of a loop at H. (c) The image of the sphere LK produced by the action of L on K. (d) Another set, homotopic to LK, in which the loop has been allowed to expand into a tube, to make clearer the relation between LK and a sphere.

$$LK(\theta,\phi)=K(2\theta,\phi),\quad 0\leq\theta\leq\pi/2;$$
$$=L(4\theta-2\pi),\quad \pi/2\leq\theta\leq\pi. \tag{2.3}$$

We can represent LK by a map Lg of the disk into G,

$$LK(\theta,\phi)=Lg(\theta,\phi)H, \tag{2.4}$$

by taking Lg to be given by

$$Lg(\theta,\phi)=g(2\theta,\phi)h_1,\quad 0\leq\theta\leq\pi/2;$$
$$=a(4\theta-2\pi)h_1^{-1}g(\pi,\phi)h_1,\quad \pi/2\leq\theta\leq\pi \tag{2.5}$$

The factors of h_1 on the right have no effect on the structure in coset space given by (2.4). Along with the factor h_1^{-1} they are there (a) to insure that the circumference of the disk is indeed taken into H_0 by the map $Lg(\pi,\phi)$ and (b) to insure that continuity is preserved at $\theta=\pi/2$.

Under the isomorphism of Theorem 2, the representative map LK of $\pi_2(G/H)$ corresponds to the representative loop $Lg(\pi,\phi)$ of $\pi_1(H_0)$. According to (2.5) this loop is just

$$Lg(\pi,\phi)=h_1^{-1}g(\pi,\phi)h_1. \tag{2.6}$$

That (2.6) does, in fact, give a loop in H_0 follows from the fact that $g(\pi,\phi)$ is such a loop and the fact that H_0 is a normal subgroup of H. It is easily verified that the homotopy class of loops in H_0 to which $h_1^{-1}g(\pi,\phi)h_1$ belongs, does not depend on the choice of h_1 within the connected component H_1 of H. Thus under the isomorphism of Theorem 1, we have associated with every element of $\pi_1(G/H)$—i.e., with each component of H—a loop $h_1^{-1}g(\pi,\phi)h_1$ in H_0 which, under the isomorphism of Theorem 2, corresponds to the homotopy class of LK in $\pi_2(G/H)$. Since $g(\pi,\phi)$ is a loop in H_0 corresponding to the homotopy class of K in $\pi_2(G/H)$, we conclude that (2.6) gives precisely the action of $\pi_1(G/H)$ on $\pi_2(G/H)$ asserted by Theorem 3.

III. ILLUSTRATIVE APPLICATIONS

A. Media that can be regarded as a field of objects with a given point group symmetry

Only proper rotations are required to relate the orientations of such objects at different points of a connected medium, and we need therefore only consider the subgroup of proper point group operations in constructing the isotropy subgroup H. Examples of such media and the proper subgroup of their point groups include nematics (D_∞), biaxial nematics (D_2), isotropic ferromagnets (C_∞), the dipole-locked A phase of superfluid helium-3 (C_0), etc.

To apply the algorithm we must take G to be the simply connected group SU(2), which is related to the multiply connected proper rotation group SO(3) through the homomorphic mapping:

$$R(\hat{n},\theta)\longleftrightarrow\pm u(\hat{n},\theta)$$
$$=\pm\exp[i(\theta/2)\hat{n}\cdot\sigma]$$
$$=\pm[\cos\tfrac{1}{2}\theta+i\sin\tfrac{1}{2}\theta\,\hat{n}\cdot\sigma],\quad 0\leq\theta\leq 2\pi, \tag{3.1}$$
$$[u(\hat{n},\theta\pm 2\pi)=-u(\hat{n},\theta)],$$

where $R(\hat{n},\theta)$ is a rotation through the angle θ about the axis $\hat{n}$, and the σ_i are the Pauli matrices. The isotropy subgroup H is then the "lift" in SU(2) of the proper part of the point group: for each proper rotation $R(\hat{n},\theta)$ in the point group, H contains the pair of 2×2 unitary matrices $\pm u(\hat{n},\theta)$ [or, alternatively, $u(\hat{n},\theta)$ and $u(\hat{n},\theta+2\pi)$]. Thus H is just the double group of the proper subgroup of the point group, familiar from its (quite unrelated) applications in the treatment of magnetic ions in crystals.

The application of the algorithm is particularly simple when the point group is discrete, for its lift in SU(2) is then also discrete, H_0 is the identity in SU(2), and H/H_0 is H itself. For a medium of objects with discrete point group symmetries the fundamental group is simply the double group associated with the proper subgroup of the point group. A particularly simple application of this is the result of Toulouse and Kleman[1] that the fundamental group of dipole-locked ^{3}He-A is Z_2, the two element group. {*Proof*: C_0 [the one element group in SO(3)] lifts to the two element group ($u=\pm1$) in SU(2).} A more interesting application is provided by Toulouse's charming observation[2] that the fundamental group of the biaxial nematic is the quaternion group Q. [*Proof*: D_2 consists of the four rotations 1, $R(\hat{x},\pi)$, $R(\hat{y},\pi)$, and $R(\hat{z},\pi)$; under the homorphism (3.1) these lift to the eight elements ±1, $\pm i\sigma_x$, $\pm i\sigma_y$, and $\pm i\sigma_z$, which *are* the quaternion group.] Evidently the procedure is similarly straightforward for fields of arbitrary objects with no continuous symmetry elements.

Point singularities are even simpler when the point group is discrete, for H_0 then consists of the identity alone, and $\pi_1(H_0)=0$. It follows that $\pi_2(G/H)=0$: there are no point defects.

To illustrate the use of the algorithm when the point group is not discrete, we consider the ordinary nematic. The proper part of the point group, D_∞, contains elements of the form $R(\hat{z},\theta)$ and $R(\hat{x},\pi)R(\hat{z},\theta)$ for arbitrary θ. These lift to the subgroup of SU(2) consisting of $u(\hat{z},\theta)$ and $i\sigma_x u(\hat{z},\theta)$ for $0\leq\theta\leq4\pi$. Thus H has two connected components. (The negative of any operation of H is in the same component as the operation itself.) H_0 is the subgroup containing the $u(\hat{z},\theta)$, and the operations $i\sigma_x u(\hat{z},\theta)$ are in the coset $H_1=(i\sigma_x)H_0$. The quotient group H/H_0 is thus the two-element group Z_2.

The topology of H_0 is that of the circle, and therefore[15] $\pi_1(H_0)=Z$, the integers. It follows from Theorem 2 that $\pi_2(G/H)=Z$ for the nematic. The automorphism group generated by π_1 consists of the single inner automorphism $H\to h^{-1}Hh$ given by any h in the component of H not connected to the identity. A simple choice is $i\sigma_x$ itself. The resulting automorphism acts on any curve $u(z,t)$ in H_0 as follows:

$$(i\sigma_x)^{-1}u(\hat{z},t)(i\sigma_x)=u(\hat{z},-t). \tag{3.2}$$

Thus a loop in the class of $\pi_1(H_0)$ characterized by winding number n is in the same automorphism class as a loop characterized by winding number $-n$. Point defects in a nematic (as emphasized by Volovik and Mineyev[7]) are thus characterized by a positive integer n indicating that they belong to the class of maps of a sphere into G/H with degree $+n$ *or* $-n$. The combination law for point defects is $n_1+n_2\to|n_1\pm n_2|$. (This has the amusing consequence that in the presence of a

line defect doubly quantized point defects are unstable: one need merely decompose a $2n$ point defect into two identical defects of type n, move one of them around the line defect to convert it to $-n$, and allow the reunited point defects to annihilate one another.)

B. Some examples from the theory of superfluid helium-3

Here the order parameter has the structure of a second-rank tensor in spin and orbital space. In general, operations in G can act differently on the two types of indices, and the simply connected group G must be taken as the group SU(2)×SU(2) consisting of ordered pairs of operations from SU(2). H is then the lift of the isotropy subgroup in SO(3)×SO(3). The elegant result of Cross and Brinkman[9] for the dipole-free A phase is recovered from the following rather bleak analysis:

The reference order parameter in the dipole-free A phase can be taken to be the tensor $A^0_{mn}=\hat{z}_m(\hat{x}_n+i\hat{y}_n)$. The isotropy subgroup in SO(3) contains operations of the form $[R(\hat{z},\theta),1]$ and $[R(\hat{x},\pi)R(\hat{z},\theta),R(\hat{z},\pi)]$. Lifting these to SU(2)×SU(2) gives a four-component subgroup consisting of

$$H_0\colon [u(\hat{z},\theta),1],\quad H_1\colon [i\sigma_x u(\hat{z},\theta),i\sigma_z],$$
$$H_2\colon [u(\hat{z},\theta),-1],\quad H_3\colon [i\sigma_x u(\hat{z},\theta),-i\sigma_z],$$
$$-2\pi\le\theta\le 2\pi. \tag{3.3}$$

Taking g to be the element

$$g=[i\sigma_x,i\sigma_z], \tag{3.4}$$

these can be represented as the following cosets of H_0:

$$H_n=g^nH_0,\quad n=0,\ldots,3. \tag{3.5}$$

The quotient group H/H_0 is therefore isomorphic to Z_4, the integers modulo 4. Theorem 1 then gives $\pi_1(G/H)=Z_4$.

Since H_0 has the topology of a circle $\pi_2(G/H)=Z$. One easily verifies that

$$g^{-n}[u(\hat{z},t),1]g^n=[u(\hat{z},(-1)^nt),1], \tag{3.6}$$

and therefore the action of π_1 on π_2 is again to group elements of $\pi_2(G/H)$ with degree n and $-n$ into a single class of defects. The import of the relations in (3.6) is that a point defect of type n must be transported around a line singularity of type g or g^3 (but not g^2) to be converted into its own inverse. As in the case of the nematic, if a g or g^3 line singularity is present it can catalyze the decay of all point singularities down to a single one of unit degree.

The more recent results of Balian and Love on the A_1 phase[10] can be similarly treated. If, for example, the reference order parameter is $A^0_{mn}=(\hat{x}_m+i\hat{y}_m)(\hat{x}_n+i\hat{y}_n)$ and the spin and orbital parts are allowed to orient arbitrarily, then the isotropy subgroup in SO(3)×SO(3) contains just those transformations of the form $[R(z,\theta),R(z,-\theta)]$. This lifts to a two-component subgroup of SU(2), consisting of

$$H_0\colon [u(\hat{z},\theta),u(\hat{z},-\theta)],\quad -2\pi\le\theta\le 2\pi,$$
$$H_1\colon [u(\hat{z},\theta),-u(\hat{z},\theta)],\quad -2\pi\le\theta\le 2\pi,$$
$$H_1=gH_0,\quad g=[1,-1]. \tag{3.7}$$

{The other pieces one might write down are already included in these families; e.g. $[-u(z,\theta),u(z,\theta)]=[u(z,\theta+\pi),-u(z,\theta+\pi)]$.} Hence H/H_0 is the two element group, and $\pi_1(G/H)=Z_2$. Once again H_0 has the topology of the circle, so $\pi_2(G/H)$ is again the integers. However $g^{-1}ag=a$ for any element of SU(2)×SU(2), so $\pi_1(G/H)$ acts trivially on $\pi_2(G/H)$: point defects with quantum numbers n and $-n$ are distinct.

IV. PROOF OF THEOREMS 1 AND 2

The proof is an application of the exact homotopy sequence:

$$\pi_2(G)\to\pi_2(G/H)\to\pi_1(H)\to$$
$$\pi_1(G)\to\pi_1(G/H)\to\pi_0(H)\to$$
$$\pi_0(G). \tag{4.1}$$

The use of this exact sequence in computing the homotopy groups of condensed matter physics has been described by Volovik and Mineyev.[7] To their analysis we need add only two remarks:

(a) As discussed in Sec. II, it is always possible to choose the group G so that $\pi_2(G)=\pi_1(G)=\pi_0(G)=0$;

(b) With this choice of G the theorems follow immediately from the general structural feature of exact sequences that a pair of groups sandwiched between a pair of trivial groups in an exact sequence must be isomorphic.

ACKNOWLEDGMENTS

I would never have learned topology without the patient help of many past and present students—particularly Bob Penner, Steve Shenker, and Jason Ho. G. Toulouse, M. Kleman, G. E. Volovik, and V. P. Mineyev have kindly kept me supplied with preprints and reprints of their very elegant expositions. It is a pleasure to thank Andrew Sommese for telling me about the remarkable theorem of Cartan, and getting me unstuck in my efforts to understand the action of π_1 on π_2. This work was supported in part by the National Science Foundation through the Materials Science Center of Cornell University, Technical Report No. 2961.

Note added in proof: Shortly after submitting this paper I received a preprint from M. Kleman and L. Michel (to be published in Phys. Rev. Lett.) which makes similar use of universal covering groups and the theorem of Cartan to compute the first and second homotopy groups of coset spaces arising in condensed matter physics.

[1] G. Toulouse and M. Kleman, J. Phys. Lett. 37, L-149 (1976).
[2] G. Toulouse, J. Phys. Lett. 38, L-67 (1977).
[3] M. Kleman, L. Michel, and G. Toulouse, J. Phys. Lett. 38, L-195 (1977).
[4] M. Kleman, J. Phys. Lett. 38, L-199 (1977).
[5] V. Poenaru and G. Toulouse, J. Phys. (Paris) 8, 887 (1977).
[6] G. E. Volovik and V. P. Mineyev, Zh. Eksp. Teor. Fiz., Pis'ma Red. 24, 605 (1976).
[7] G. E. Volovik and V. P. Mineyev, Zh. Eksp. Teor. Fiz. 72, 2256 (1977).

[8]V. P. Mineyev and G. E. Volovik, Zh. Eksp. Teor. Fiz. **73**, 767 (1977); see also "Planar and Linear Solitons in Superfluid He-3", preprint submitted to Phys. Rev. B in August 1977.

[9]M. C. Cross and W. F. Brinkman, J. Low Temp. Phys. **27**, 683 (1977).

[10]D. Balian and A. Love (preprint).

[11]For a possible exception see V. P. Mineyev, G. E. Volovik, and N. D. Mermin, "Topological Analysis of the Cores of Singularities in ^{3}He-A", submitted to J. Low Temp. Physics.

[12]An unusually lucid expoistion of this and many other important topological properties of continuous groups can be found in L. Pontrjagin, *Topological Groups* (Princeton U. P., Princeton, N.J., 1939).

[13]A proof is given by H. Borel, in *Seminaire Henri Cartan* 1949/1950 (Benjamin, New York, 1967), article No. 12.

[14]Proof that the subgroup H_0 is a normal subgroup: Let g belong to H, and a belong to H_0. Since H_0 is the connected component of the identity there is a continuous family of elements in H_0, $a(t)$, joining e to a. The family $g^{-1}a(t)g$ is also continuous, but joins e to $g^{-1}ag$. Hence $g^{-1}ag$ is also in the connected component of the identity. (The proof that H_0 is a subgroup runs along similar lines.)

[15]This can also be proved by an application of Theorem 1. A point on the circumference of a circle can be specified by giving the member of SO(2) that produces it by action on a fixed reference point. The isotropy subgroup is then the identity of SO(2). The simply connected covering group G of SO(2) is the one-dimensional translation group, the homomorphism being $T(\theta + 2\pi n) \to R(\theta)$, $n = 0, \pm 1, \pm 2, \cdots$. The isotropy subgroup H is the discrete subgroup of translations $T(2\pi n)$. Since H is discrete $H/H_0 = H$, and π_1 is isomorphic to H itself, which in turn is isomorphic to the additive group of the integers.

An inversionlike integral equation in the multidimensional case

Henri Cornille

Rockefeller University, New York, New York 10021
and D.P.H.T. Saclay, BP n° 2, 91190 Gif-sur-Yvette, France
(Received 7 July 1977)

In this paper and in the following, we deduce inversionlike integral equations (from which we can construct a class of potentials without introducing the data) in the case of a system of n linear first-order partial differential equations with different coordinate variables. The partial differential part of the linear operator is a diagonal matrix with each element acting only on a particular coordinate.

I. INTRODUCTION

The connection, using the so-called "inverse Method" between one-dimensional nonlinear partial differential equations (nlpde) and the associated linear one-dimensional coordinate differential systems is actually relatively well understood.[1] The link is obtained by considering the one dimensional inversion equation associated to this linear system. However, the problem in the multidimensional case[2] is not in as good shape as in the one-dimensional case, because the inversion corresponding to a multidimensional partial differential system is missing.[3]

Let us consider $D_0(x) = \partial/\partial x$ and

$$(\Delta_0(x_1, x_2, \ldots, x_n) + ik\Lambda - Q)\,\psi = 0, \tag{1}$$

where Λ is a diagonal eigenvalue matrix $\Lambda = (\delta_{ij}\lambda_i)$, Q a $(n \times n)$ "potential"

$$Q = \begin{pmatrix} q_1^1 \cdots q_1^n \\ \\ q_n^1 \cdots q_n^n \end{pmatrix},$$

and ψ a column vector. Δ_0 is a diagonal partial differential operator $\Delta_0 = (\delta_{ij} D_0(x_i))$.

Let us consider a set of solutions $[D_0(x) + ik\lambda_i]U_i^0(x) = 0$ and define a set of eigenfunctions of Δ_0,

$$\psi_j^0 = (\delta_{i,j} U_j^0(x_j)), \quad (\Delta_0 + ik\Lambda)(\psi_1^0, \ldots, \psi_n^0) = (0).$$

Let us formally write a set of n functions [which we would like to be solutions of Eq. (1)] with the following representation,

$$\psi_j = (\delta_{ij} U_j^0(x_j) + \int_{x_j}^{\infty} U_j^0(y) G_i^j(x_1, x_2, \ldots, x_n; y)\,dy). \tag{2}$$

Our aim (as it was in the one coordinate case[4]) in this paper and in the following[5] is to obtain, from algebraic methods, integral equations (IE) for the transforms G_i^j such that we can generate both a class of solution ψ_j [Eq. (2)] and of potential Q of the system (1). We shall not try to define (as it must be for a true inverse formalism) the data on the basis of which Q could be reconstructed, nor to get from these data the kernel of the IE. Thus we want to provide a potential framework for an inversionlike procedure. Let us recall that in the one-dimensional case,[4] this procedure is useful to generate the solutions of the nlpde associated with linear differential systems as well as to construct a class of energy dependent (or angular momentum dependent) potentials.

By changing the variables, (1) can be enlarged to more general systems. Let us start with a system (1a) where $\Delta_0 = [\delta_{ij}\mu_i(x_i)(\partial/\partial x_i)]$, with $\mu_i > 0$, $\lim_{x\to-\infty} \int^x (\mu_i(u))^{-1} du = +\infty$. Equation (2) becomes (2a) $\psi_j = [\delta_{ij} U_j^0(x_j) + \int_{x_j}^{\infty} \mu_j^{-1} G_i^j dy)$ with now $(\mu_j(x)\partial/\partial x + i\lambda_j)U_j^0 = 0$. However by a change of variables $\bar{x}_i = \int^{x_i} (\mu_i(u))^{-1} du$ we come back to (1) with $\Delta_0 = (\delta_{ij}\partial/\partial \bar{x}_i)$.

In Sec. II, subsections A, B, and C we formally derive an IE corresponding to Eq. (1) when any two elements of Δ_0 apply to two different coordinates, and in subsection F when two, three, $\cdots$ of these elements apply to the same coordinate. In subsection D we discuss the conditions to be satisfied by the kernel of the IE in order that all the formalism be correct, and in subsection E the properties of the solutions of our IE.

II. INVERSIONLIKE INTEGRAL EQUATIONS

A. Statement of the problem

First we write down a representation of a set of solutions of the whole system as transforms of the reduced set of solutions when the "potential" is switched off. Secondly, we put this representation into the differential system and we get (boundary conditions appear at this stage) that these transforms must satisfy well defined nlpde. Thirdly, we guess an integral equation (IE) such that the solutions satisfy the above nlpde. In this way we get that the kernels of the IE must satisfy well-defined partial differential equations and also boundary conditions in order that all the derivations be correct. A drawback of the method is that we have no guarantee concerning the uniqueness of the IE and we shall see in the following paper[5] that for (1)–(2) different IE can exist, i.e., can lead to the same above set of nlpde.

B. nlpde satisfied by $G_i^j(x_1, \ldots, x_n; y)$

We assume the following boundary conditions,

$$\lim_{s\to-\infty} U_i^0(s) G_j^i(x_1, x_2, \ldots, x_n; s) = 0, \quad j \neq i \text{ and } j = i. \tag{3}$$

Let us introduce the following Lemma:

 0022-2488/78/1906-1463$1.00

Lemma: If the boundary conditions (3) are satisfied, then we get:

$$[D_0(x_i)+ik\lambda_i]\,[U_i^0(x_i)+\int_{x_i}^{\infty}U_i^0G_i^i\,ds]$$
$$=\int_{x_i}^{\infty}ds\;U_i^0(D_0(x_i)+D_0(s))G_i^i$$
$$[D_0(x_j)+ik\lambda_j]\Big(\int_{x_i}^{\infty}U_i^0G_j^i\,ds\Big)$$
$$=\frac{\lambda_j}{\lambda_i}U_i^0(x_i)G_j^i(\cdots,x_i)$$
$$+\int_{x_i}^{\infty}U_i^0\Big(D_0(x_j)+\frac{\lambda_j}{\lambda_i}D_0(s)\Big)G_j^i\,ds. \tag{4}$$

We assume that the ψ_j given by (2) are solutions of (1),

$$(\Delta_0+ik\Lambda_i-Q)(\psi_1,\psi_2,\ldots,\psi_n)=(0), \tag{5}$$

and we define

$$O^i_{j,x_j}=D_0(x_j)+\frac{\lambda_j}{\lambda_i}D_0(y),\quad \hat{G}^i_j=G^i_j(x_1,\ldots,x_n;x_i).$$

We assume that the conditions of the lemma are satisfied, and taking into account Eq. (4), the result Eq. (5) is written down with scalar quantities

$$-q_i^iU_i^0(x_i)+\int_{x_i}U_i^0(y)(O^i_{i,x_i}G^i_i-\textstyle\sum q_i^mG^i_m)\,dy=0,$$
$$\Big(-q_j^i+\frac{\lambda_j}{\lambda_i}\hat{G}^i_j\Big)U_i^0(x_i)+\int_{x_i}U_i^0(y)(O^i_{j,x_j}G^i_j-\textstyle\sum q_j^mG^i_m)\,dy=0,\quad i\neq j. \tag{5'}$$

It follows that if the transforms $\{G_j^i\}$ of Eq. (2) satisfy the nlpde,

$$O^i_{jx_j}G^i_j=\sum_{m\neq j}G^i_m\hat{G}^m_j\frac{\lambda_j}{\lambda_m}, \tag{6}$$

then the set of potentials $\{q_j^i\}$ of Eq. (1)

$$q_j^i=\frac{\lambda_j}{\lambda_i}\hat{G}^i_j,\quad q_i^i=0, \tag{7}$$

are such that the lhs of Eq. (5) is zero.

C. A conjecture of an inversion like integral equation of Eq. (1)

Let us consider the following integral equation: $F_i^i=0$ and

$$K_j^i(x_1,\ldots,x_n;y)=F_j^i(x_j,y)+\sum_{m\neq i}^{m=n}\int_{x_m}^{\infty}F_m^i(s,y)\times K_j^m(x_1,\ldots,x_n;s)\,ds. \tag{8}$$

For the $\{F_j^i\}$ and the $\{K_j^i\}$ we assume the following boundary condition,

$$\lim_{y\to\infty}F_j^i(x,y)=0,$$
$$\lim_{j\to\infty}F_j^i(s,y)K_i^j(x_1,x_2,\cdots,x_n;s)=0 \tag{9}$$

and we require of course that the solution of Eq. (8) exists.

Property: If we assume that each kernel F_j^i satisfies both the boundary condition Eq. (9) and the partial differential equation

$$O^i_{jx_j}F^i_j=\Big(D_0(x_j)+\frac{\lambda_j}{\lambda_i}D_0(y)\Big)F^i_j(x_j,y)=0, \tag{10}$$

then the solution of Eq. (8) satisfies the nlpde [Eq. (6)]

$$O^i_{jx_j}K^i_j=\sum_{m\neq j}\frac{\lambda_j}{\lambda_m}K^i_m\hat{K}^m_j,\quad \hat{K}^m_j=K^m_j(x_1,\ldots,x_n;y=x_m). \tag{6a}$$

For the proof we apply O^i_{j,x_j} to both side of Eq. (8),

$$O^i_{jx_j}K^i_j=O^{ii}_{jx_j}F^i_j-F^i_j\hat{K}^j_j+\sum_m\int F^i_mD_0(x_j)K^m_j$$
$$+\sum_m\int K^m_j\frac{\lambda_j}{\lambda_i}D_0(y)F^i_m\,ds. \tag{11}$$

Using the relation (10), the rhs of Eq. (11) can be written

$$\sum_{m\neq j}\frac{\lambda_j}{\lambda_m}F^i_m\hat{K}^m_j+\sum_m\int F^i_mO^m_{jx_j}K^m_j\,ds.$$

and comparing with the solution of Eq. (8), the result (6′) follows. Let us define $\mathcal{K}(x_1,\ldots,x_n;y)=(K^i_j(x_1,\ldots,x_n;y))$ and

$$\mathcal{J}(x_1,\ldots,x_n;y)=(F^i_j(x_j,y)),$$
$$\tilde{\mathcal{J}}(s,y)=(F^i_j(s,y)\;\theta(s-x_j)). \tag{12}$$

then Eq. (8) can be written in matrix form

$$\mathcal{K}(x_1,\ldots,x_n;y)=\mathcal{J}(x_1,\ldots,x_n;y)+\int_{-\infty}^{+\infty}\tilde{\mathcal{J}}(s,y)\times\mathcal{K}(x_1,\ldots,x_n;s)\,ds, \tag{8a}$$

where in the jth column of $\mathcal{J}$ only the x_j coordinate appears.

In conclusion, if the kernel $\mathcal{J}$ satisfies both Eqs. (9) and (10), if we substitute the solution $\mathcal{K}$ of Eq. (8) into the representation (2), $K^i_j=G^i_j$, and further if the condition (3) is satisfied, then Eq. (2) are solutions of our starting partial differential system (1) and consequently Eq. (8) will be an associated IE.

D. Conditions on the kernel

The discussion is the same as in the one-dimensional case,[4] the conditions being at infinity in both cases. We must verify (9) and that the solution of the IE [Eq. (8)] exists.

$\lambda_i\lambda_j<0$: We can have square integrable scalar kernels with a continuum as well as a discrete part,

$$F^i_j(x_j,y)=\int d\nu^i_j\alpha^i_j(\nu^i_j)\exp[i\nu^i_j(\lambda_jx_j-\lambda_iy)], \tag{13a}$$

or

$$F^i_j(x_j,y)=\sum_m\alpha^i_{j,m}\exp[-\nu^i_{j,m}(x_j-\lambda_i\lambda_j^{-1}y)],\quad \nu^i_{j,m}>0. \tag{13b}$$

$\lambda_i\lambda_j>0$: For simplicity we consider discrete kernels: They are not square integrable,

$$F^i_j(x_j,y)=\sum_m\alpha^i_{j,m}\exp\nu^i_{j,m}(x_j-\lambda_i\lambda_j^{-1}y),\quad \nu^i_{j,m}>0. \tag{14}$$

Whereas in (13b), $F^i_j(x_j,y)\to 0$ either when $x_j\to\infty$ or $y\to\infty$, on the contrary in (14), $F^i_j(x_j,y)\underset{y\to\infty}{\to}0$ and they behaved badly when $x_j\to\infty$: $F^i_j(x_j,y)\to\infty$, $x_j\to\infty$.

Let us call $\mathcal{J}^+(\mathcal{J}^-)$ the set of (F^i_j) kernels with $\lambda_i\lambda_j>0$ ($\lambda_i\lambda_j<0$). With $\mathcal{J}^-$ are associated square integrable scalar kernels $F^i_j(x_j,y)$, whereas with $\mathcal{J}^+$ are associated badly behaving kernels as in Eq. (14).

If we assume that the kernels are discrete and

degenerate of the most simple form $F_j^i(x_j,y)=g_j^i(x_j)h_j^i(y)$ given by (15b) and (14), then from simple examples[4] we learn for $\mathcal{J}^*$ that we cannot have both F_j^i and F_i^j, $F_j^iF_k^jF_i^k$ all present, and more generally $F_{i_1}^{i_2}F_{i_2}^{i_3}\cdots F_{i_n}^{i_{n-1}}F_{i_1}^{i_n}$ all present.

We rewrite the IE, Eq. (8a) in the form

$$K(x_1,\ldots,x_n;y)=\sum_0^\infty K_n(x_1,x_2,\ldots,x_n;y),\quad K_0=\mathcal{J},\tag{15}$$

$$\begin{aligned}&K_1=\int_{-\infty}^{+\infty}\tilde{\mathcal{J}}(s,y)\mathcal{J}(x_1,\ldots,x_n;s)\,ds,\\&(K_1)_j^i=\int_{-\infty}^{+\infty}\Big(\sum_l F_l^i(s,y)F_j^l(x_j,s)\theta(s-x_l)\Big)ds.\end{aligned}\tag{16}$$

When $s\to-\infty$ we must require that all these terms be integrable

$$\forall(i,j),\quad \sum_l F_l^i(s,y)F_j^l(x_j,s)\underset{s\to-\infty}{\to}0,\tag{17}$$

sufficiently quickly in order that the integral exists and we must check the compatibility of these conditions.

In this way, we again find the above rules for $F_j^i\in\mathcal{J}^*$. Let us consider

$$\begin{aligned}&\int_{-\infty}^{+\infty}\tilde{\mathcal{J}}(s,v)\tilde{\mathcal{J}}(w,s)\,ds\\&\quad=\int_{-\infty}^{+\infty}\Big(\sum_l F_l^i(s,y)F_j^l(w,s)\theta(s-x_l)\theta(w-x_j)\Big)ds\end{aligned}\tag{18}$$

for v and w fixed. We remark that for degenerate kernels, as those considered in this section, the integrability conditions (13b) and (14) when $s\to-\infty$ are both the same in Eq. (16) and Eq. (18). K_n can be written

$$\begin{aligned}K_n=&\int\tilde{\mathcal{J}}(u_1,y)\tilde{\mathcal{J}}(u_2,u_1)du_1\cdots\int\tilde{\mathcal{J}}(u_{i+1},u_i)\\&\times\tilde{\mathcal{J}}(u_{i+2},u_{i+1})\,du_{i+1}\cdots\int du_n\tilde{\mathcal{J}}(u_n,u_{n-1})\mathcal{J}(x_1,\ldots,x_n;u_n).\end{aligned}$$

We see that the same integrability condition Eq. (17) or Eq. (18) is repeated in the general case and so we have no new condition in order that K_n exists. We remark that our choice of the y dependence of the kernels F_j^i in Eq. (13b) and Eq. (14) is such that $\lim_{y\to-\infty}\tilde{\mathcal{J}}(s,y)=0$, further y appears in $K_n(x_1,\ldots,x_n;y)$ only from the first matrix $\tilde{\mathcal{J}}(u_1,y)$. It follows that $K_n(x_1,\ldots,x_n;y)\to0$ when $y\to0$. Similarly the behavior of $\tilde{\mathcal{J}}(s,y)K_n(x_1,x_2,\ldots,x_n;s)$ when s is large is given by $\tilde{\mathcal{J}}(s,y)\mathcal{J}(u_1,s)$ and so goes to zero when $s\to\infty$.

E. Properties of the solutions of the IE Eq. (8)

1. We shall see that the potentials K_j^i and the solutions K_j^i satisfy besides the nlpde (6a) [necessary in order that Eq. (8) be an IE] other nlpde. For any $k\neq j$, $D_0(x_k)F_j^i(x_j,y)=D_0(x_k)F_l^i(s,y)=0$ and consequently from Eq. (8) we get:

$$D_0(x_k)K_j^i=-F_k^i\hat{K}_j^k+\sum_l\int_{x_l}F_l^iD_0(x_k)K_j^l\tag{19}$$

$$\begin{aligned}&\frac{\partial}{\partial x_k}K_j^i=-K_k^i\hat{K}_j^k,\quad k\neq j,\quad n\geq2,\\&\frac{\partial}{\partial x_k}\hat{K}_j^i=-\hat{K}_k^i\hat{K}_j^k,\quad k\neq i\neq j\neq k,\quad n\geq3.\end{aligned}\tag{20}$$

2. Due to $[D_0(u)+(\lambda_j/\lambda_i)D_0(v)]F_j^i(u,v)=0$, then $F_j^i(u,v)=F_j^i(\lambda_ju-\lambda_jv)$. We shall see that consequently the potentials $\hat{K}_j^i$ constructed from Eq. (8) depend on only $n-1$ independent variables. Equation (8) can be written

$$\begin{aligned}K_j^i(x_1,\ldots,x_n;y)=&F_j^i(\lambda_jx_j-\lambda_iy)\\&+\sum_l\int_{x_l}^\infty F_l^i(\lambda_ls-\lambda_iy)K_j^l(x_1,\ldots,x_n;s)\,ds.\end{aligned}\tag{21}$$

Let us define $\tilde{K}_j^i(z)=K_j^i(x_1,\ldots,x_n;\,y=x_i+z)$, $x_{ji}=\lambda_jx_j-\lambda_ix_i$. We have $x_{ji}+x_{ij}=0$, $x_{ij}+x_{jk}+x_{ki}=0$. Equation (8) can be rewritten

$$\tilde{K}_j^i(z)=F_j^i(x_{ji}-\lambda_iz)+\sum_l\int_0^\infty F_l^i(x_{li}+\lambda_lu-\lambda_iz)\tilde{K}_j^l(u)\,du.\tag{22}$$

Then $\tilde{K}_j^i(z)$ depends on z, $\{x_{ij}\}$ whereas $\tilde{K}_j^i(z=0)=\hat{K}_j^i$ depend on the set $\{x_{ij}\}$, $i=1,\ldots,n$, $j=1,\ldots,n$, $i\neq j$. Only $(n-1)$ independent x_{ij} exist. Take for instance the set $x_{12},x_{13},\ldots,x_{1n}$, then for any other (i,j) $x_{ij}=x_{1j}-x_{1i}$. Consequently $\hat{K}_j^i$ depends at most on $(n-1)$ variables. For $n=2$ we have only the variable x_{12}; for $n=3$, two variables x_{12}, x_{13}; and so on.

F. Inversion equations in some particular less than n coordinate case

Let us assume in (1) that the first q coordinates are different, $x_1\neq x_2\neq\cdots\neq x_q$ while the $(n-q)$ remaining ones are equal, $x_q=x_{q+1}=\cdots=x_n$,

$$[\Delta_0(x_1,x_2,\ldots,x_q,x_q,\ldots,x_q)+ik\Lambda-Q]\,\psi=0.\tag{1'}$$

The elements of the diagonal Δ_0 matrix are $D_0(x_1)$, $D_0(x_2),\ldots,D_0(x_q)$, $D_0(x_q),\ldots,D_0(x_q)$ whereas Λ is the same eigenvalue matrix as in (1).

Let us consider the representation

$$\psi_j=(\delta_{i,j}U_j^0(x_j)+\int_{x_j}^\infty U_j^0(y)G_i^j(x_1,x_2,\ldots,x_q,x_q,\ldots,x_q;y)dy)\tag{2'}$$

with $x_j=x_q$ for $j\geq q$. For $j\leq q$, the definition of ψ_j is the same as in Eq. (2) whereas for $j>q$, the ψ_j have their integration path starting at x_q.

We assume the boundary condition Eq. (3) with $x_j=x_q$ for $j>q$ and call it condition (3'). The first important difference in the formalism (apart from the trivial change $x_j=x_q$ for $j>q$) appears in the second relation of the lemma (Eq. 4) when both $i>q$, $j>q$, $i\neq q$ $[(\lambda_j/\lambda_i-1)U_i^0\hat{G}_j^i$ instead on $\lambda_j/\lambda_iU_i^0\hat{G}_j^i]$. If we require that the representation (2') is a solution of (1'), if we assume (3'), and apply (4'), then we get an Eq. (5'), like Eq. (5) with two changes:

(i) A trivial one where x_j for $j>q$ must be replaced by x_q.

(ii) A nontrivial one due to the supplementary term in Eq. (4') for both $i\geq q$, $j\geq q$, $i\neq j$. This leads in the lhs of Eq. (5) to a supplementary term, $-U_i^0(x_q)\hat{G}_j^i$.

From Eq. (5'), it follows that if

$$\begin{aligned}&O_{j,x_j}^iG_j^i=\sum_{m\neq j}G_m^i\hat{G}_j^m\gamma_j^m,\\&\gamma_j^m=\begin{cases}\left(\dfrac{\lambda_j}{\lambda_m}-1\right),&j\geq q\text{ and }m\geq q,\\\lambda_j(\lambda_m)^{-1},&\text{either }j<q\text{ or }m<q,\end{cases}\end{aligned}\tag{6'}$$

then

$$q_i^j = 0, \quad q_i^j = \begin{cases} \lambda_i(\lambda_j)^{-1}\hat{G}_i^j, & \text{either } j < q \text{ or } i < q, \\ \left(\frac{\lambda_i}{\lambda_j} - 1\right)\hat{G}_i^j, & i \geq q \text{ and } j \geq q. \end{cases} \tag{7'}$$

Let us consider Eq. (8) when $x_j = x_q$ for $j \geq q$, $F_i^j = 0$,

$$K_j^i(x_1, \ldots, x_q, \ldots, x_q; y) = F_j^i(x_j, y) + \sum_{m \neq i} \int_{x_m}^{\infty} F_m^i(s, y) K_j^m(x_1, \ldots, x_q; s)\, ds. \tag{8'}$$

Let us assume both the boundary conditions (9) with $x_i = x_q$ for $i > q$ [and call it (9')] and the partial differential equations

$$\left(D_0(x_i) + \frac{\lambda_i}{\lambda_j} D_0(y)\right) F_i^j(x_i, y) = 0, \quad x_i = x_q \text{ for } i \geq q; \tag{10'}$$

then we show that the set $\{K_j^i, \hat{K}_j^i\}$, solutions of Eq. (8') satisfies the nlpde [Eq. (6')].

$$O_{j, x_j}^i K_j^i = \sum_{m \neq j} \gamma_j^m K_m^i \hat{K}_j^m. \tag{6'a}$$

For the proof we apply $O_{jx_j}^i$ to both sides of Eq. (8') and we get that

$$O_{jx_j}^i K_j^i - \sum_m \int F_m^i O_{jx_j}^m K_j^m$$

equals

$$\sum_{m \neq j} \frac{\lambda_j}{\lambda_m} F_m^i \hat{K}_j^m \quad \text{if } j < q,$$

$$\sum_{m=1}^{q-1} \frac{\lambda_j}{\lambda_m} F_m^i \hat{K}_j^m + \sum_{\substack{m=q \\ m \neq j}}^{n} F_n^i \left(1 - \frac{\lambda_j}{\lambda_m}\right) \hat{K}_j^m \quad \text{if } j \geq q.$$

Comparing with the IE (8'), the result (6'a) follows.

The sufficient conditions on F_j^i in order that all the derivations be correct are the same in both formalisms Eqs. (1)–(8) or Eqs. (1')–(8'). For the properties of the solutions of the IE Eq. (8') we get that the potentials $\hat{K}_j^i$ depend on $(q-1)$ independent variables $(x_{12}, x_{13}, \ldots, x_{1q})$, and the K_j^i, $\hat{K}_j^i$ have to satisfy nlpde Eq. (20') [like Eq. (20) with $x_l = x_q$ for $l \geq q$].

III. CONCLUSION

In our algebraic method, the derivation of the IE associated with a linear differential (or partial differential) system is done in two successive steps. First we get the nlpde that the transforms of the solutions of the system must satisfy. Secondly we seek an IE such that the solutions satisfy these nlpde. However, the IE exhibited here are such that their kernels $F_j^i(s, y)\theta(s - x_j)$ are almost independent of the coordinates, whereas the free terms $F_j^i(x_j, y)$ depend on only one coordinate. Further they satisfy lpde such that they depend in fact on only one variable. We find that the reconstructed potentials satisfy *extra nlpde Eq.* (20) and we show explicitly that they depend in fact on only $n-1$ variables (linked to the n coordinates). Moreover the degenerated kernels are of the *pure exponential type* and so the values of the corresponding *potentials cannot be confined* in the more than one coordinate space. These results suggest that the potentials reconstructed from the present IE Eq. (8) represent a subclass of the whole class of potentials associated with the system (1).

In the following paper[5] the kernels F_j^i will depend on all the coordinates, satisfy lpde different of Eq. (10), and we shall show that there exists a generalization of the present IE [the solutions still satisfying the nlpde Eq. (6a)] differing only if the number of different coordinates is higher than one. This means that in the multidimensional inversion formalism entirely new features not present in the one-dimensional case appear. For instance,[5] the kernels of this other IE *are not necessarily of the pure exponential type*; there exist *potentials confined* in a two-dimensional space *which do not satisfy extra nlpde.*

In conclusion the structure of the multidimensional inversion formalism appears very rich and complex and this paper is a first investigation in order to explore it.

In their first theory Zakharov and Shabat[2] considered an IE corresponding to one coordinate and introduced the other as parameters. Although it is not at first sight obvious that this is identical to the determination of the IE associated to a linear partial differential system, we think that their formalism and the one presented here and in Ref. 5 must have some analogy. (This point will be discussed elsewhere).

[1] For a review paper see, for instance, A. C. Scott *et al.*, I. E. E. E. **61**, 1443 (1973).

[2] V. E. Zakharov and A. B. Shabat, Func. Anal. Appl. **8**, 226 (1974). *Note added*: After the completion of this work we have become aware that these authors have obtained new results in the two-dimensional case: S. P. Novikov, talk at the "International Conference on the Mathematical Problems in Theoretical Physics" (Rome, June 1977).

[3] M. J. Ablowitz and R. Haberman, P. R. L. **35**, 1185 (1975).

[4] H. Cornille, J. Math. Phys. **17**, 2143 (1976); **18**, 1855 (1977); Rockefeller Preprint C00-2232B-124 (to be published in J. Math. Phys.).

[5] H. Cornille: "Confined solutions of multidimensional inversion equations," (to be published in J. Phys. A).

Asymptotic steady-state solution in the functional random-walk model

I. Hosokawa

Department of Mechanical Engineering, Iwate University, Morioka 020, Japan
(Received 22 November 1977)

The maximum-entropy (equilibrium) state of the functional random-walk model of a closed many-particle system is known, but there has been no direct proof that this state represents the steady solution of the basic functional equation in the model. Here, a direct proof of it is given for the asymptotic case where the average particle-number density is extremely large.

1. INTRODUCTION

In the previous works,[1-3] the dynamics of the functional random-walk model was established so as to simulate closely the Liouville dynamics for a classical many-particle system in the sense of a kind of coarse graining. The model develops in an irreversible way from any initial state towards the maximum-entropy state, which corresponds to the equilibrium state in statistical mechanics.

In the functional random-walk model, we deal with, rather than $6N$-dimensional phase points, the *stochastic* particle-number density field in the one-body phase space. The equation for the probability density functional of the field is written in the Fokker—Planck or Kolmogorov form, which guarantees the irreversible development of the system only if there exists any interaction between particles. This equation has the same content as the functional equation of Bogoliubov[4] and the author,[5] which is equivalent to the Liouville dynamics in the thermodynamic limit, except for the one essential modification leading to the irreversibility; various physical discussions were made on the coarse-graining nature of this modification, which was introduced rather from a mathematical motivation.[1-3] The basic Fokker— Planck or Kolmogorov functional equation in our model dynamics is explained in the next section.

The maximum-entropy state of the probability density functional was easy to obtain in our model dynamics,[2,3] and it was deduced at this state that the most probable form of the stochastic particle-number density field is canonical. However, as for the question whether it represents a steady solution of the basic functional equation, we have no direct answer beyond a pretty reasonable, but indirect supporting argument. It is because the basic functional equation contains a complicated operator with the modification described above. In this paper, we formally give the direct answer to this problem for the asymptotic case where the average particle-number density n is exceedingly greater than unity. Of course, n depends on the physical scale of length. Then, our asymptotic state may always be realized by taking a big scale of length. Therefore, the present direct proof of the asymptotic steady-state solution will remain effective on most of many-particle systems; until another direct proof for an arbitrary n appears in the future.

2. BASIC DYNAMICS OF THE FUNCTIONAL RANDOM-WALK MODEL

The basic Fokker—Planck or Kolmogorov equation may be written as

$$\frac{\partial\hat{\rho}\delta[B]}{\partial t}=-\int_X\frac{\delta}{\delta z(x)}\{Qz(x)\hat{\rho}\delta[B]\}dx + \frac{1}{2}\int_X\int_X\frac{\delta^2}{\delta z(x)\delta z(x')}\{D(z;x,x')\hat{\rho}\delta[B]\}dxdx', \tag{2.1}$$

which governs the probability density functional $\hat{\rho}$ of the particle-number density field $nz(x)$ in the one-body phase space $(x\in)X$. Here, t is the time variable, $\delta/\delta z(x)$ the functional derivative with respect to z at x, Q the nonlinear Vlasov operator defined as

$$Qz(x)=[H_1(x);z(x)]+n\int_X[\phi(q-q');z(x)z(x')]dx' \tag{2.2}$$

in which $H_1(x)$ is the one-body Hamiltonian, $[\,;\,]$ the Poisson bracket, and $\phi(q)$ the interaction potential between particles, q being the displacement vector, and

$$D(z;x,x')\equiv P[\phi(q-q');z(x)z(x')] \tag{2.3}$$

where P is the operator which makes the subsequent symmetric function positive definite (see Ref. 1). $\delta[B]$ associated with $\hat{\rho}$ indicates that the probability density $\hat{\rho}(z)$ in the function space A of z is meaningful only inside its subspace B; in other words, $\delta[B]\delta z$ is the measure in A which vanishes outside B but is the Riemannian measure inside B. This measure forces the normal component of all $z(x)$ to the space B to be trivial or vanishing in (2.1). The space B is defined by the following conditions:

$$\lim_{V\to\infty}\int_X z(x)dx/V=1 \tag{2.4}$$

and

$$\lim_{V\to\infty}\Big\{\int_X nH_1(x)z(x)dx + \int_X\int_X\tfrac{1}{2}n^2\phi(|q-q'|)z(x)z(x')dxdx'\Big\}/V=\text{const}, \tag{2.5}$$

where V is the volume containing the whole system. The former equation prescribes the normalization of z and the latter the average energy of the system in question. As (2.5) is nonlinear, B is Riemannian

 0022-2488/78/1906-1467$1.00

in A. Furthermore, the condition $z(x) \geq 0$ is necessary for the physical meaning of z to hold.

The expression of the $\hat{\rho}$ equation associated with $\delta[B]$ was first presented in Ref. 3. This expression is valid, against the original one in Refs. 1 and 2, in the case when $z(x)$ is considered as a Euclidean vector in A. We assumed in these works that $z(x)$ could represent a set of Riemannian coordinates in B, but this is unjustifiable. Therefore, the present form of the equation was adopted. [The Fokker–Planck or Kolmogorov equation in the Riemannian space is usually well treated in the frame of curvilinear coordinates. But there is some advantage for physical understanding in dealing with the present form equation, however curious it seems to be at first glance.] However, all the results in Refs. 1 and 2 will remain essentially valid, if we proceed with the present equation. Only a formal change is necessary in the process, such as replacing $\hat{\rho}$ by $\hat{\rho}\delta[B]$; accordingly we have the equilibrium measure $\hat{\rho}_\infty \delta[B]\delta z$ in A in place of $\hat{\rho}_\infty \delta z$ (assumedly in B). $\hat{\rho}_\infty$ represents the maximum-entropy state, which is explicitly given as

$$\hat{\rho}_\infty = A_\infty \exp(-n\int_X z \log z\, dx), \tag{2.6}$$

where A_∞ is the constant to be determined by the time-invariant normalization condition:

$$\int_A \hat{\rho}(z, t)\delta[B]\delta z = 1. \tag{2.7}$$

It may cause some question in mathematics to deal with a probability density in infinite dimensions in the above-described formal way. In order to evade this difficulty, it is reasonable to understand the whole formalism in a finite-dimensional approach to $z(x)$ (the cylinder-functional approach[1]) and to take a limit in the final formulas, assuming the convergence.

Finally, it is notable that the Liouville dynamics in the thermodynamic limit is completely recovered if we take away the P operator from the D factor as well as the factor $\delta[B]$ from the whole equation (2.1).[1] The resulting equation is nothing but the functional Fourier transform of the Bogoliubov–Hosokawa equation[4,5] for the generating or state functional, which is equivalent to the so-called BBGKY hierarchy. The irreversibility in question as well as the physical interpretation of $\hat{\rho}$ as the probability density of the particle-number density field came into play only with the modification by P.[1–3]

3. THE ROLE OF THE OPERATOR P

A general solution for $\hat{\rho}(z, t)\delta[B]$ may be given in terms of the functional Green kernel G. That is,

$$\hat{\rho}(z, t)\delta[B] = \int_A G(z, t/z', 0)\hat{\rho}(z', 0)\delta[B]\delta z'; \tag{3.1}$$

$$G(z, t/z', t') = \lim_{\Delta t \to 0} \int_A \cdots \int_A P_{\Delta t}(z^L/z^{L-1}) \cdots P_{\Delta t}(z^1/z^0) \times \prod_{k=1}^{L-1} \delta z^k, \tag{3.2}$$

where $z = z^L$, $z' = z^0$, $\Delta t = (t - t')/L$, and the superscripts indicate the order of time subintervals. The infinitesimal Green kernel $P_{\Delta t}$ is explicitly obtained from (2.1) as

$$P_{\Delta t}(z^{k+1}/z^k) = \int_A \exp\{i\int_X y^k(x)[z^k(x) - z^{k+1}(x) + \Delta t Q z^k(x)]dx - \tfrac{1}{2}\Delta t \int_X \int_X y^k(x) y^k(x') D(z^k; x, x') dx dx'\}\delta y^k \delta[B]^{k+1}, \tag{3.3}$$

where $\delta[B]^k$ denotes $\delta[B]$ for z^k and $y^k(x) \in A$. Obviously, $P_{\Delta t}(z^{k+1}/z^k)$ is Gaussian for z^{k+1} in B, whose parameters depend on z^k. Hence, it may be said that (3.2) with (3.3) provides a generalized kind of the Wiener measure. G involves the probability of all possible paths of $z(x)$ which start from z' at t' and reach z at t in the discretized form; this may be called the path sum according to Dekker.[6]

In fact, functional integration in (3.3) is performed to yield the explicit Gaussian form in z^{k+1}, as follows. Since D is a symmetric kernel, there must be the one orthogonal transformation in A, which makes diagonal the quadratic form in y^k in the exponent in (3.3) for a definite z^k. After this transformation, we have the new function $y^{k*} = T_T y^k$ in place of y^k and the new eigenvalue function $D^*(z^k; x)\delta(x - x')$ of $D(z^k; x, x')$. Here, T is the functional version of the transformation matrix. Using a scalar-product expression,[7] we may write

$$\int_X y^k(z^k - z^{k+1} + \Delta t Q z^k)dx = \langle T y^{k*},\ z^k - z^{k+1} + \Delta t Q z^k\rangle = \langle y^{k*},\ {}^tT(z^k - z^{k+1} + \Delta t Q z^k)\rangle, \tag{3.4}$$

$$\int_X \int_X y^k(x) D(z^k; x, x') y^k(x') dx dx' = \langle T y^{k*},\ D T y^{k*}\rangle = \langle y^{k*},\ {}^tT D T y^{k*}\rangle = \langle y^{k*},\ (D^*\delta) y^{k*}\rangle = \int_X (y^{k*})^2 D^*(z^k; x) dx. \tag{3.5}$$

Thus, we have

$$\begin{aligned} P_{\Delta t}(z^{k+1}/z^k) &= \int_A \exp\{i\int_X y^{k*}[{}^tT(z^k - z^{k+1} + \Delta t Q z^k)]dx \\ &\quad - \tfrac{1}{2}\Delta t \int_X (y^{k*})^2 D^*(z^k; x) dx\}\delta y^{k*}\delta[B]^{k+1} \\ &= \exp\{-\tfrac{1}{2}\int_X dx [{}^tT(z^k - z^{k+1} + \Delta t Q z^k)]^2 / [\Delta t D^*(z^k; x)]\} \\ &\quad \times \prod_j [\Delta t D^*(z^k; x_j)]^{-1/2} \delta[B]^{k+1}, \end{aligned} \tag{3.6}$$

noting that $\delta y = \prod_j dy(x_j)(\Delta x_j / 2\pi)^{1/2}$, by definition,[5] and $D^* \geq 0$. When $D^*(z^k; x) = 0$, the exponential factor acts as a delta function only for that x and z^k.

It is important to note here that

$${}^tT[\phi(|q - q'|); z(x)z(x')]T = D^*\delta + N^*\delta \tag{3.7}$$

by definition,[2] where $N^*\delta$ is the negative definite part of the eigenvalue function of $[\phi; zz]$. The P operator causes to neglect $N^*\delta$, so that we have

$$D^*\delta = U({}^tT[\phi; zz]T)\delta\, {}^tT[\phi; zz]T, \tag{3.8}$$

where $U(a)$ is the step function such that it is unity for $a \geq 0$, and vanishes for $a < 0$. Since $U\delta$ as well as ${}^tT[\phi; zz]T$ are diagonal and then commute with each other, we may write

$$D^*\delta = U\delta\, {}^tT[\phi; zz]T = {}^tT[\phi; zz]T U\delta. \tag{3.9}$$

Hence

$$D = TD^*\delta^t T = TU\delta^t T[\phi; zz] = [\phi; zz]TU\delta^t T. \quad (3.10)$$

This may be rewritten as

$$D(z; x, x') = \int_X u(z; x, x'')[\phi(|q'' - q'|); z(x'')z(x')]dx''. \quad (3.11)$$

Obviously, the function u is a functional of z and also symmetric with respect to interchange of the arguments.

Furthermore, we give some comments on the quality of T and u. We have, by virtue of orthogonality,

$$\delta = T^t T, \quad \delta(x - x') = \int_X T(x, x'')T(x', x'')dx'', \quad (3.12)$$

which is equal to

$$\delta(x' - x) = \int_X T(-x, x'')T(-x', x'')dx''. \quad (3.13)$$

As a result, we have

$$T(x, x'') = \pm T(-x, x''); \quad (3.14)$$

in other words, T as a function of the first argument x should be either even or odd, depending on the second argument x'' as a parameter. Since x is a six-dimensional vector $\equiv (q_1, q_2, q_3, p_1, p_2, p_3)$ where $p \equiv (p_1, p_2, p_3)$ denotes the momentum vector of a particle, the above-described parity of the function T holds for each component of x independently. Remember

$$\delta(x - x') \equiv \prod_{i=1}^{3} \delta(q_i - q_i')\delta(p_i - p_i'). \quad (3.15)$$

Then, u may be expressed as

$$u(z; x, x') = \int_X \int_X T(x, x'')\tilde{U}(z; x'')\delta(x'' - x''')T(x', x''')dx''dx'''$$

$$= \int_X T(x, x'')T(x', x'')\tilde{U}(z; x'')dx'', \quad (3.16)$$

where $\tilde{U}$ is identical with U but re-expressed as a function in X. This shows that, by virtue of the parity of T, u can be divided into the two parts of the integral, the one of which is even in x_μ (any component of x) and the other odd in x_μ. As is evident from the form of (3.16), the part even in x_μ should be at the same time even in x'_μ, while the part odd in x_μ be also odd in x'_μ. We will call these parts of the integral (3.16) the x_μ-even u part and the x_μ-odd u part, respectively.

4. PROOF OF THE ASYMPTOTIC STEADY SOLUTION

First, we note that $\hat{\rho}_\infty$ in (2.6) has the saddle point in B, which is determined by

$$\delta[-n\int_X z\log z dx] = 0 \quad (4.1)$$

under the conditions (2.4) and (2.5). That is,[2]

$$z_B(x) = \exp\{-\lambda - \beta[H_1(x) + n\int_X \phi(|q - q'|)z_B(x')dx']\}, \quad (4.2)$$

λ and β being the constants to be determined by those two conditions. It is easily known that z_B satisfies

$$Qz_B(x) = 0. \quad (4.3)$$

$\hat{\rho}_\infty\delta[B]$ can be expressed, as follows, in the neighborhood of this saddle point;

$$\hat{\rho}_\infty\delta[B] = A_\infty \exp\left[-n\int_X z_B\log z_B dx - \frac{n}{2}\int_X \frac{1}{z_B}(z - z_B)^2 dx\right] \times \delta[B]. \quad (4.4)$$

Here, if n is extremely large, the probability measure $\hat{\rho}_\infty\delta[B]\delta z$ tends to vanish for all z, except for $z = z_B$ where it tends to be concentrated to be unity.

With this fact in mind, let us insert $\hat{\rho}_\infty$ into (2.1) for $\hat{\rho}$. Then, the first term in the right-hand side is

$$-\int_X \frac{\delta}{\delta z(x)}\{Qz\hat{\rho}_\infty\delta[B]\}dx = n\int_X Qz_B(\log z_B + 1)dx\hat{\rho}_\infty\delta[B]$$

$$= 0 \quad (4.5)$$

on account of (4.3). The second term is

$$\frac{1}{2}\int_X\int_X \frac{\delta^2}{\delta z(x)\delta z(x')}\{D(z; x, x')\hat{\rho}_\infty\delta[B]\}dxdx'$$

$$= \tfrac{1}{2}\int_X\int_X D(z_B; x, x')\{n^2[\log z_B(x) + 1][\log z_B(x') + 1]$$

$$- \frac{n}{z_B(x)}\delta(x - x')\}dxdx'\hat{\rho}_\infty\delta[B]. \quad (4.6)$$

With the aid of (3.11) and (4.2), the right-hand side of (4.6) becomes

$$\tfrac{1}{2}\int_X\int_X\int_X u(z_B; x, x'')\frac{\partial\phi}{\partial q''}(|q'' - q'|)\beta\left(\frac{\partial H_1}{\partial p'} - \frac{\partial H_1}{\partial p''}\right)z_B(x'')z_B(x')$$

$$\times\left\{n^2[\log z_B(x) + 1][\log z_B(x') + 1]\right.$$

$$\left. - \frac{n}{z_B(x)}\delta(x - x')\right\}dxdx'dx''\hat{\rho}_\infty\delta[B].$$

In the integral with the first term in the curly bracket, the part with $\partial H_1/\partial p'$ vanishes because $\partial H_1/\partial p'$ is odd in the momentum space while the other factors relevant to p' are all even in it, as is seen in (4.2). Next, consider the part with $\partial H_1/\partial p''$. Since $\partial H_1/\partial p''$ is odd in the momentum space, only the p''-odd u part of (3.16) contributes to the integral. But it is necessarily odd in p, as was noted upon (3.16). Therefore, the part vanishes with the integral with respect to p, because the remaining factor is even in p. Thus, we are left with the integral with the second term in the curly bracket that is rewritten as

$$-\frac{n}{2}\int_X\int_X u(z_B; x, x'')\frac{\partial\phi}{\partial q''}(|q'' - q|)\beta$$

$$\times\left(\frac{\partial H_1}{\partial p} - \frac{\partial H_1}{\partial p''}\right)z_B(x'')dxdx''\hat{\rho}_\infty\delta[B]$$

in which both parts with $\partial H_1/\partial p$ and $\partial H_1/\partial p''$ must vanish for the same reason as described above in the last argument on the part with $\partial H_1/\partial p''$.[8] In consequence, the whole term of (4.6), and then all the terms in the right-hand side of (2.1), should vanish for $\hat{\rho}\delta[B] = \hat{\rho}_\infty\delta[B]$ for the asymptotic case, $n \to \infty$.

This concludes the proof. It is apparent that the asymptotic steady solution $\hat{\rho}_\infty\delta[B]$ is unique and stable, from the argument that the entropy defined on the basis of $\hat{\rho}\delta[B]$ should continue to increase until $\hat{\rho}_\infty\delta[B]$ is reached.[2]

[1]I. Hosokawa, J. Math. Phys. **11**, 657 (1970).
[2]I. Hosokawa, J. Math. Phys. **14**, 1374 (1973); Prog. Theor. Phys. **52**, 1513 (1974).
[3]I. Hosokawa, to be published in the Proceedings of NATO Advanced Study Institute 1977, Antwerpen, on Path Integrals and Their Applications in Quantum. Statistical and Solid State Physics.
[4]N.N. Bogoliubov, in *Studies in Statistical Mechanics*, edited by J. de Boer and G.E. Uhlenbeck (North-Holland, Amsterdam, 1962), Vol. 1.
[5]I. Hosokawa, J. Math. Phys. 8, 221 (1967).
[6]H. Dekker, Physica A 85, 363 and 598 (1976).
[7]B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1965).
[8]In the above argument, it was assumed that the kinetic energy part of H_1 does not include any electromagnetic vector potential. However, even with it the same argument can apply if we know that (3.15) equals

$$\delta(x-x')=\prod_{i=1}^{3}\delta(q_i-q_i')\delta\{[p_i-eA_i^e(q)/c]-[p_i'-eA_i^e(q')/c]\},$$ so

that the parity of T holds for the new translated vector $\equiv(q_1, q_2, q_3, p_1-eA_1^e/c, p_2-eA_2^e/c, p_3-eA_3^e/c)$. Here, e is the electronic charge, c the light speed, A^e the electromagnetic vector potential.

Lorentz subgroup analysis of the Lie algebra of Sp(8,*R*) and a null-plane, boson realization

L. P. Staunton

Department of Physics and Astronomy, Drake University, Des Moines, Iowa 50311
(Received 16 January 1978)

An algorithm is given for the construction of the Lie algebra of Sp(2^N, R), integral $N \geq 3$, in the form in which the tensor character of the generators under the action of a Lorentz subgroup is apparent. A specific realization of the algebra of Sp(8,R) in terms of four boson operators is presented, along with the tensor form of the complete algebra classified under the (unique) full null-plane Lorentz subalgebra. The identities obtaining for the specific boson operator realization employed for Sp(8,R) are listed, use of which permits subsidiary conditions for four boson constrained systems to be expressed in terms of Lorentz invariants.

1. INTRODUCTION

It is well known[1] that the Lie algebra of Sp($2n$,R) may be given a concrete, self-adjoint realization in terms of the $n(2n+1)$ bosonic operators consisting of all independent bi-linear combinations of n variables q_j, and their canonical conjugates $\eta_j \equiv -i\partial/\partial q_j$, $j=1,2,\ldots,n$. As is the case with any Lie algebra, different linear combinations of the generators may be selected so as to present the algebra in some standard, convenient form, e.g., the Cartan form, which is useful for general classification purposes, construction of Dynkin diagrams, etc.

For the purpose of certain physical applications, however, the most convenient form of a Lie algebra is one in which the generators are arranged into combinations which transform as tensors under a particular Lorentz subgroup. In the case of the orthogonal groups, the general technique for effecting an arrangement of the generators into such a tensor form is simple and well known. Unfortunately, due to the higher level of complexity of their root diagrams,[1] no general technique of equal simplicity has been given for large symplectic groups, and those Lorentz classifications which have been reported are not systematic.[2]

In this communication, the results for the Lie algebra $\mathcal{L}$Sp(8,R) are presented in detail, along with an inductive algorithm applicable to $\mathcal{L}$ Sp(2^N,R), for integral $N>3$.

In the simpler cases, $N=1$ and $N=2$, the Lie algebras obtained are isomorphic to those of SU(1,1) and SO(3,2), respectively. The latter isomorphism with an orthogonal group is particularly convenient: the 10 generators of Sp(4,R) can be immediately classified into a vector and a second rank tensor under a Lorentz subgroup, and a realization of these generators in terms of bilinears in q_i, η_j $(i,j=1,2)$ which yields the Majorana representation[3] of the Lorentz group has been given by Dirac.[2] No convenient isomorphisms obtain[1] for $N \geq 3$.

Historically, however, there has been no compelling reason to inquire for the Lorentz tensor character of the generators of higher order symplectic groups, nor to seek concrete bosonic realizations of the generators. The advent of the dual resonance model,[4] and more particularly, the emergence of the model known as the Nambu string,[5] however, have recently focused attention upon physical systems whose algebraic structure is composed of bilinear boson operators. The Lorentz classification of such operators has therefore become a problem of immediate interest, and constitutes the motivation for this work.

The 1973 effort at quantization of the Nambu string model by Goddard, Goldstone, Rebbi, and Thorn,[6] who employed an analysis in terms of null-plane boson operators, yielded the unphysical result that the postulated generators of the Poincaré group achieved the required algebraic closure only in 26 space–time dimensions. Subequently, Marnelius[7] developed a manifestly Lorentz convariant analysis of the string in terms of the classical Dirac-bracket procedure for constrained Hamiltonian systems, and concluded that difficulties with the generators of translations were apparent, even at the classical level. Further, he found that no covariant, bosonic-normal mode degree of freedom could be canonical, and that, as a consequence, ordering problems at the level of first quantization would be doubly severe in a covariant formalism (as opposed to a null-plane formalism).

In a parallel line of development, a spin-$\frac{1}{2}$ positive-energy relativistic wave equation was presented[8] whose structure was based upon operators comprising the two boson realization of Sp(4,R) ≡ SO(3,2). It was shown[9] that this model described, at the classical (non-quantum) limit level, a relativistic extended system composed of two permanently bound null-plane constituents[10] which orbit one another at the velocity of light, i.e., the classical Nambu string model restricted to excitation of only the lowest covariant normal mode.

Consequently, the covariant operator commutation relations following from the Heisenberg picture analysis of this model were obtained.[11] The results, which described a 4-space quantized Nambu string containing only first normal mode operators, were concluded to comprise the nucleus of the complete quantized string relations obtaining in 4 dimensions.[12]

The results presented here are the necessary mathematical basis for an extension of the Sp(4,R) wave equation model to a system containing several covariant

 0022-2488/78/1906-1471$1.00

normal modes, i.e., a system containing bilinears in the null-plane operators q_i, η_j, for $i, j = 1,2,3,4$. Further communications will discuss wave equations and Heisenberg models constructed from these operators. In particular, the question of the proper form of commutation relations between operators describing covariant string normal modes of different order is now under investigation.

In Sec. 2, a general result due to Browne[13] is adapted to form part of an algorithm for sequential classification of the generators of $\mathcal{L}\mathrm{Sp}(2^N,R)$. In Sec. 3, the particular results obtained for $\mathrm{Sp}(8,R)$ are detailed, and in Sec. 4, the special operator identities which obtain for the realization of $\mathcal{L}\mathrm{Sp}(8,R)$ in terms of boson operators are cataloged.

2. AN ALGORITHM BASED UPON A REAL CLIFFORD ALGEBRA

Consider the following set of real 4×4 matrices realizing a five-dimensional (Dirac-type) Clifford algebra with metric $(-1, 1, -1, 1, 1)$:

$$\gamma_1=\begin{bmatrix}0 & \sigma_1\\ -\sigma_1 & 0\end{bmatrix},\quad \gamma_2=\begin{bmatrix}0 & i\sigma_2\\ -i\sigma_2 & 0\end{bmatrix},$$
$$\gamma_3=\begin{bmatrix}0 & \sigma_3\\ -\sigma_3 & 0\end{bmatrix},\quad \gamma_4=\begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix},\qquad (2.1)$$
$$\gamma_5=\begin{bmatrix}0 & I\\ I & 0\end{bmatrix}.$$

As is well known, a complete, linearly independent set of 16 (real) 4×4 matrices may be simply obtained from those above. Let this set be partitioned into the set $\{S\}$ of 10 symmetric matrices, and the set $\{A\}$ of 6 antisymmetric matrices. Then

$$\{S\}=\{I,\gamma_2,\gamma_4,\gamma_5,\gamma_1\gamma_2,\gamma_1\gamma_4,\gamma_1\gamma_5,\gamma_2\gamma_3,\gamma_3\gamma_4,\gamma_3\gamma_5\},$$
$$\{A\}=\{\gamma_1,\gamma_3,\gamma_1\gamma_3,\gamma_2\gamma_4,\gamma_2\gamma_5,\gamma_4\gamma_5\},\qquad (2.2)$$

and with these particular choices, simple closure of both commutation and anticommutation relations between any two matrices is immediate.

Let $i, j = 1,2,\ldots,10$ number the elements of $\{S\}$, and $k, l = 1,2,\ldots,6$ number those of $\{A\}$, and define real 8×8 matrices as follows, in an obvious notation:

$$\Gamma_i(S,\text{ even},+)=\begin{bmatrix}S_i & 0\\ 0 & S_i\end{bmatrix},$$
$$\Gamma_i(S,\text{ even},-)=\begin{bmatrix}S_i & 0\\ 0 & -S_i\end{bmatrix},$$
$$\Gamma_i(S,\text{ odd},+)=\begin{bmatrix}0 & S_i\\ S_i & 0\end{bmatrix},\qquad (2.3)$$
$$\Gamma_k(A,\text{ odd},-)=\begin{bmatrix}0 & A_k\\ -A_k & 0\end{bmatrix}.$$

Clearly, the sets of matrices Γ exhibited above constitute a choice and partition of the linearly independent elements M_A, $A = 1,\ldots36$ of the set of *all* 8×8 matrices $\{M\}$ of the form

$$M=\begin{bmatrix}S & Z\\ Z^T & \hat{S}\end{bmatrix},\qquad (2.4)$$

where S and $\hat{S}$ are *any* two 4×4 real symmetric matrices, and Z is *any* 4×4 real matrix ($Z^T \equiv$ transpose of Z). The convenience of this particular partition will become clear below.

Let the 8×8 matrix β be defined as

$$\beta=\begin{bmatrix}0 & I\\ -I & 0\end{bmatrix},\qquad (2.5)$$

from which the properties $\beta^T=-\beta$, $\beta^2=-I$ follow, and consider the set of all linearly independent matrices $\hat{M}_A$, $A = 1,2,\ldots,36$ obtained from those of the set $\{M_A\}$ via

$$\hat{M}_A=\beta M_A.\qquad (2.6)$$

Now the complete set of *all* such 8×8 matrices has the general form

$$\hat{M}=\begin{bmatrix}Z^T & \hat{S}\\ -S & -Z\end{bmatrix}.\qquad (2.7)$$

It follows, then, from their form (2.7) alone,[1] that *the 36 linearly independent matrices $\hat{M}_A$ constitute a defining 8×8 matrix realization of the Lie algebra of* $\mathrm{Sp}(8,R)$.

Consider next the set of 4 real variables $\{q_1,q_2,q_3,q_4\}$ and their canonical operator conjugates $\{\eta_1,\eta_2,\eta_3,\eta_4\}$ ($\eta \equiv -i\partial/\partial q$). Define the column matrix

$$Q \equiv \mathrm{col}(q_1,q_2,q_3,q_4,\eta_1,\eta_2,\eta_3,\eta_4),\qquad (2.8)$$

and denote the entries Q_a, $a = 1,2,\ldots,8$. The result

$$[Q_a,Q_b]=i\beta_{ab},\quad a,b=1,2,\ldots,8,\qquad (2.9)$$

where β_{ab} are the elements of the matrix β of (2.5) follows immediately.

Now the result of Browne[13] states, in this notation, that if there exists a set of matrices $\{\hat{M}_A\}$ with the property

$$\beta\hat{M}_A{}^T\beta=\hat{M}_A,\quad \text{all } A,\qquad (2.10)$$

which satisfy any Lie algebra, then the quantities

$$G_A=\tfrac{1}{4}Q^TM_AQ,\quad \text{all } A,\qquad (2.11)$$

[where the M_A are obtained from the $\hat{M}_A$ via (2.6)] satisfy the same Lie algebra. Since the matrices $\hat{M}_A$ exhibited above satisfy (2.10) and the Lie algebra of $\mathrm{Sp}(8,R)$, and since the matrices M_A, and in particular the partitioned sets $\{\Gamma\}$ of (2.3), are obtained via (2.6), it follows that the 36 operators G defined, generically, as

$$\{G\} \equiv \tfrac{1}{4}Q^T\{\Gamma\}Q\qquad (2.12)$$

form a bosonic representation of the Lie algebra $\mathcal{L}\mathrm{Sp}(8,R)$.

The particular form of the generators G defined via the algorithm (2.2), (2.3) is singularly convenient since, in this case, the commutator of any two operators G is equal to the factor $\pm i$ times another *single* G operator. In this form, the Lorentz tensor character of the operators may be immediately identified. *Let the symbol ~ denote equality up to sign and the numerical factors of $\frac{1}{4}$ appearing in* (2.12), *and let the operators G be denoted below by their particular Γ content.* One then obtains,[14] for example, in a self-explanatory notation:

$$[\Gamma_i(S,\text{even},+),\Gamma_j(S,\text{even},+)]\sim\begin{bmatrix}0 & [S_i,S_j]\\ -[S_i,S_j] & 0\end{bmatrix}$$
$$\equiv \Gamma([S_i,S_j],\text{odd},-)$$
$$[\Gamma_i(S,\text{even},+),\Gamma_j(S,\text{even},-)]\sim\begin{bmatrix}0 & \{S_i,S_j\}\\ \{S_i,S_j\} & 0\end{bmatrix} \quad (2.13)$$
$$\equiv\Gamma(\{S_i,S_j\},\text{odd},+).$$

(The full compelement of relations is exhibited in Table I.) *The convenience of the result depends, then, upon the simple closure of both commutation and anticommutation relations of the sets of* 4×4 *matrices* (2.2).

The 36 operators G and their Lie algebra, $\mathcal{L}\,\mathrm{Sp}(8,R)$, classified according to their tensor character under a null-plane Lorentz subgroup will be exhibited in the following section. Here we continue with brief remarks concerning the extension of the algorithm to $\mathcal{L}\,\mathrm{Sp}(2^N,R)$, for integral $N>3$.

To the set of 36 matrices Γ may be appended a second set of 28 matrices $\hat{\Gamma}$ of the form

$$\hat{\Gamma}_k(A,\text{even},+)=\begin{bmatrix}A_k & 0\\ 0 & A_k\end{bmatrix},$$
$$\hat{\Gamma}_k(A,\text{even},-)=\begin{bmatrix}A_k & 0\\ 0 & -A_k\end{bmatrix}, \quad (2.14)$$
$$\hat{\Gamma}_i(S,\text{odd},-)=\begin{bmatrix}0 & S_i\\ -S_i & 0\end{bmatrix},$$
$$\hat{\Gamma}_k(A,\text{odd},+)=\begin{bmatrix}0 & A_k\\ A_k & 0\end{bmatrix},$$

TABLE I. Format of commutation relation results between members of the partitioned sets of generators G, denoted here by their Γ matrix content.

$\Gamma_i(S,\text{even},+)$,	$\Gamma_j(S,\text{even},+)\sim\Gamma([S_i,S_j],\text{odd},-)$
$\Gamma_i(S,\text{even},+)$,	$\Gamma_j(S,\text{even},-)\sim\Gamma(\{S_i,S_j\},\text{odd},+)$
$\Gamma_i(S,\text{even},+)$,	$\Gamma_j(S,\text{odd},+)\sim\Gamma(\{S_i,S_j\},\text{even},-)$
$\Gamma_i(S,\text{even},+)$,	$\Gamma_k(A,\text{odd},-)\sim\Gamma([S_i,A_k],\text{even},+)$
$\Gamma_i(S,\text{even},-)$,	$\Gamma_j(S,\text{even},-)\sim\Gamma([S_i,S_j],\text{odd},-)$
$\Gamma_i(S,\text{even},-)$,	$\Gamma_j(S,\text{odd},+)\sim\Gamma(\{S_i,S_j\},\text{even},+)$
$\Gamma_i(S,\text{even},-)$,	$\Gamma_k(A,\text{odd},-)\sim\Gamma([S_i,A_k],\text{even},-)$
$\Gamma_i(S,\text{odd},+)$,	$\Gamma_j(S,\text{odd},+)\sim\Gamma([S_i,S_j],\text{odd},-)$
$\Gamma_i(S,\text{odd},+)$,	$\Gamma_k(A,\text{odd},-)\sim\Gamma([S_i,A_k],\text{odd},+)$
$\Gamma_k(A,\text{odd},-)$,	$\Gamma_l(A,\text{odd},-)\sim\Gamma([A_k,A_l],\text{odd},-)$

so that the union of the two sets constitutes a particular choice of form of the complete set of 64 linearly independent real 8×8 matrices. Moreover, *this complete set of matrices exhibits simple closure under both commutation and anticommutation operations*. Therefore, the 64 elements of this set may be re-partitioned according to their symmetry or antisymmetry and used in place of the sets of 4×4 matrices $\{S\}$ and $\{A\}$ of Eq. (2.2). An exactly similar procedure to that carried out above for $\mathcal{L}\,\mathrm{Sp}(8,R)$ may now be followed to obtain a bosonic representation of $\mathcal{L}\,\mathrm{Sp}(16,R)$ in a form in which the Lorentz tensor character of the generators may be obtained by inspection. Clearly, the process can be continued, stepwise, to any $\mathcal{L}\,\mathrm{Sp}(2^N,R)$, $N>3$.

3. THE NULL-PLANE LORENTZ CLASSIFICATION OF $\mathcal{L}\,\mathrm{SP}(8,R)$

In order to effect a tensor classification of the generators G constructed according to (2.11) from the sets of matrices Γ of (2.3), it is necessary to select one of the Lorentz subalgebras contained in $\mathcal{L}\,\mathrm{Sp}(8,R)$, and to identify the six generators involved. For our purposes, the simple generalization to four variables of the null-plane Lorentz subalgebra of $\mathrm{Sp}(4,R)$ employed by Dirac[15] suffices.

Let the antisymmetric tensor operator $S_{\mu\nu}$ ($\mu,\nu=0,1,2,3$) generate the Lorentz subgroup. We identify, then,[16]

$$S_{10}=\tfrac{1}{4}(q_1^2-q_2^2+q_3^2-q_4^2-\eta_1^2+\eta_2^2-\eta_3^2+\eta_4^2),$$
$$S_{20}=\tfrac{1}{2}(-q_1q_2-q_3q_4+\eta_1\eta_2+\eta_3\eta_4),$$
$$S_{30}=\tfrac{1}{2}(q_1\eta_1+q_2\eta_2+q_3\eta_3+q_4\eta_4-2i), \quad (3.1)$$
$$S_{12}=\tfrac{1}{2}(q_1\eta_2-q_2\eta_1+q_3\eta_4-q_4\eta_3),$$
$$S_{23}=\tfrac{1}{2}(-q_1q_2-q_3q_4-\eta_1\eta_2-\eta_3\eta_4),$$
$$S_{31}=\tfrac{1}{4}(-q_1^2+q_2^2-q_3^2+q_4^2-\eta_1^2+\eta_2^2-\eta_3^2+\eta_4^2).$$

The designation of this particular Lorentz subalgebra as the null-plane algebra follows[15,17] from the absence of terms quadratic in η in the null-plane combinations S_{30}, S_{12}, $S_{10}+S_{13}$, and $S_{20}+S_{23}$. After lengthy inspection of the full set of commutation relations, the remaining 30 self-adjoint generators may be classified into two antisymmetric tensors, $E_{\mu\nu}$ and $F_{\mu\nu}$, four vectors A_μ, B_μ, C_μ and D_μ and two scalars T_1 and T_2. The specific operator forms and the associated Γ matrix for each component are given in Table II. The complete Lie algebra is exhibited below[18]:

$$[S_{\mu\nu},S_{\alpha\beta}]=i(g_{\mu\alpha}S_{\nu\beta}-g_{\nu\alpha}S_{\mu\beta}+g_{\nu\beta}S_{\mu\alpha}-g_{\mu\beta}S_{\nu\alpha}),$$
$$[S_{\mu\nu},E_{\alpha\beta}]=i(g_{\mu\alpha}E_{\nu\beta}-g_{\nu\alpha}E_{\mu\beta}+g_{\nu\beta}E_{\mu\alpha}-g_{\mu\beta}E_{\nu\alpha}),$$
$$[S_{\mu\nu},F_{\alpha\beta}]=i(g_{\mu\alpha}F_{\nu\beta}-g_{\nu\alpha}F_{\mu\beta}+g_{\nu\beta}F_{\mu\alpha}-g_{\mu\beta}F_{\nu\alpha}),$$
$$[S_{\mu\nu},A_\alpha]=i(g_{\mu\alpha}A_\nu-g_{\nu\alpha}A_\mu),$$
$$[S_{\mu\nu},B_\alpha]=i(g_{\mu\alpha}B_\nu-g_{\nu\alpha}B_\mu),$$
$$[S_{\mu\nu},C_\alpha]=i(g_{\mu\alpha}C_\nu-g_{\nu\alpha}C_\mu),$$

TABLE II. Specific realization of the generators of Sp(8, R) in terms of bosonic operators, and the associated Γ matrix.

$S_{10}=\frac{1}{4}(q_1^2-q_2^2+q_3^2-q_4^2-\eta_1^2+\eta_2^2-\eta_3^2+\eta_4^2)$	$\Gamma(\gamma_1\gamma_2,$ even, $-)$
$S_{20}=\frac{1}{2}(-q_1q_2-q_3q_4+\eta_1\eta_2+\eta_3\eta_4)$	$\Gamma(-\gamma_2\gamma_3,$ even, $-)$
$S_{30}=\frac{1}{2}(q_1\eta_1+q_2\eta_2+q_3\eta_3+q_4\eta_4-2i)$	$\Gamma(I,$ odd, $+)$
$S_{12}=\frac{1}{2}(q_1\eta_2-q_2\eta_1+q_3\eta_4-q_4\eta_3)$	$\Gamma(\gamma_1\gamma_3,$ odd, $-)$
$S_{23}=\frac{1}{2}(-q_1q_2-q_3q_4-\eta_1\eta_2-\eta_3\eta_4)$	$\Gamma(-\gamma_2\gamma_3,$ even, $+)$
$S_{31}=\frac{1}{4}(-q_1^2+q_2^2-q_3^2+q_4^2-\eta_1^2+\eta_2^2-\eta_3^2+\eta_4^2)$	$\Gamma(-\gamma_1\gamma_2,$ even, $+)$
$T_1=\frac{1}{2}(-q_1\eta_4+q_2\eta_3-q_3\eta_2-q_4\eta_1)$	$\Gamma(-\gamma_2,$ odd, $+)$
$T_2=\frac{1}{2}(q_1\eta_3+q_2\eta_4-q_3\eta_1-q_4\eta_2)$	$\Gamma(\gamma_4\gamma_5,$ odd, $-)$
$A_0=\frac{1}{4}(q_1^2+q_2^2+q_3^2+q_4^2+\eta_1^2+\eta_2^2+\eta_3^2+\eta_4^2)$	$\Gamma(I,$ even, $+)$
$A_1=\frac{1}{2}(-q_1\eta_1+q_2\eta_2-q_3\eta_3+q_4\eta_4)$	$\Gamma(-\gamma_1\gamma_2,$ odd, $+)$
$A_2=\frac{1}{2}(q_1\eta_2+q_2\eta_1+q_3\eta_4+q_4\eta_3)$	$\Gamma(\gamma_2\gamma_3,$ odd, $+)$
$A_3=\frac{1}{4}(q_1^2+q_2^2+q_3^2+q_4^2-\eta_1^2-\eta_2^2-\eta_3^2-\eta_4^2)$	$\Gamma(I,$ even, $-)$
$B_0=\frac{1}{2}(q_1q_4-q_2q_3-\eta_1\eta_4+\eta_2\eta_3)$	$\Gamma(\gamma_2,$ even, $-)$
$B_1=\frac{1}{2}(q_1\eta_4+q_2\eta_3-q_3\eta_2-q_4\eta_1)$	$\Gamma(\gamma_1,$ odd, $-)$
$B_2=\frac{1}{2}(q_1\eta_3-q_2\eta_4-q_3\eta_1+q_4\eta_2)$	$\Gamma(\gamma_3,$ odd, $-)$
$B_3=\frac{1}{2}(q_1q_4-q_2q_3+\eta_1\eta_4-\eta_2\eta_3)$	$\Gamma(\gamma_2,$ even, $+)$
$C_0=\frac{1}{4}(q_1^2+q_2^2-q_3^2-q_4^2+\eta_1^2+\eta_2^2-\eta_3^2-\eta_4^2)$	$\Gamma(\gamma_4,$ even, $+)$
$C_1=\frac{1}{2}(-q_1\eta_1+q_2\eta_2+q_3\eta_3-q_4\eta_4)$	$\Gamma(-\gamma_3\gamma_5,$ odd, $+)$
$C_2=\frac{1}{2}(q_1\eta_2+q_2\eta_1-q_3\eta_4-q_4\eta_3)$	$\Gamma(\gamma_1\gamma_5,$ odd, $+)$
$C_3=\frac{1}{4}(q_1^2+q_2^2-q_3^2-q_4^2-\eta_1^2-\eta_2^2+\eta_3^2+\eta_4^2)$	$\Gamma(\gamma_4,$ even, $-)$
$D_0=\frac{1}{2}(q_1q_3+q_2q_4+\eta_1\eta_3+\eta_2\eta_4)$	$\Gamma(\gamma_5,$ even, $+)$
$D_1=\frac{1}{2}(-q_1\eta_3+q_2\eta_4-q_3\eta_1+q_4\eta_2)$	$\Gamma(\gamma_3\gamma_4,$ odd, $+)$
$D_2=\frac{1}{2}(q_1\eta_4+q_2\eta_3+q_3\eta_2+q_4\eta_1)$	$\Gamma(-\gamma_1\gamma_4,$ odd, $+)$
$D_3=\frac{1}{2}(q_1q_3+q_2q_4-\eta_1\eta_3-\eta_2\eta_4)$	$\Gamma(\gamma_5,$ even, $-)$
$E_{10}=\frac{1}{4}(q_1^2-q_2^2-q_3^2+q_4^2-\eta_1^2+\eta_2^2+\eta_3^2-\eta_4^2)$	$\Gamma(\gamma_3\gamma_5,$ even, $-)$
$E_{20}=\frac{1}{2}(-q_1q_2+q_3q_4+\eta_1\eta_2-\eta_3\eta_4)$	$\Gamma(-\gamma_1\gamma_5,$ even, $-)$
$E_{30}=\frac{1}{2}(q_1\eta_1+q_2\eta_2-q_3\eta_3-q_4\eta_4)$	$\Gamma(\gamma_4,$ odd, $+)$
$E_{12}=\frac{1}{2}(q_1\eta_2-q_2\eta_1-q_3\eta_4+q_4\eta_1)$	$\Gamma(\gamma_2\gamma_5,$ odd, $-)$
$E_{23}=\frac{1}{2}(-q_1q_2+q_3q_4-\eta_1\eta_2+\eta_3\eta_4)$	$\Gamma(-\gamma_1\gamma_5,$ even, $+)$
$E_{31}=\frac{1}{4}(-q_1^2+q_2^2+q_3^2-q_4^2-\eta_1^2+\eta_2^2+\eta_3^2-\eta_4^2)$	$\Gamma(-\gamma_3\gamma_5,$ even, $+)$
$F_{10}=\frac{1}{2}(-q_1q_4-q_2q_3-\eta_1\eta_4-\eta_2\eta_3)$	$\Gamma(\gamma_1\gamma_4,$ even, $+)$
$F_{20}=\frac{1}{2}(-q_1q_3+q_2q_4-\eta_1\eta_3+\eta_2\eta_4)$	$\Gamma(\gamma_3\gamma_4,$ even, $+)$
$F_{30}=\frac{1}{2}(q_1\eta_4-q_2\eta_3+q_3\eta_2-q_4\eta_1)$	$\Gamma(-\gamma_2\gamma_4,$ odd, $-)$
$F_{12}=\frac{1}{2}(-q_1\eta_3-q_2\eta_4-q_3\eta_1-q_4\eta_2)$	$\Gamma(-\gamma_5,$ odd, $+)$
$F_{23}=\frac{1}{2}(-q_1q_3+q_2q_4+\eta_1\eta_3-\eta_2\eta_4)$	$\Gamma(\gamma_3\gamma_4,$ even, $-)$
$F_{31}=\frac{1}{2}(q_1q_4+q_2q_3-\eta_1\eta_4-\eta_2\eta_3)$	$\Gamma(-\gamma_1\gamma_4,$ even, $-)$

$$
\begin{aligned}
&[S_{\mu\nu},D_\alpha]=i(g_{\mu\alpha}D_\nu-g_{\nu\alpha}D_\mu),\\
&[S_{\mu\nu},T_1]=0,\quad [S_{\mu\nu},T_2]=0,\\
&[T_1,A_\mu]=iB_\mu,\quad [T_1,B_\mu]=iA_\mu,\\
&[T_1,C_\mu]=0,\quad [T_1,D_\mu]=0,\\
&[T_1,T_2]=0,\quad [T_1,E_{\mu\nu}]=-iF_{\mu\nu},\quad [T_1,F_{\mu\nu}]=-iE_{\mu\nu},\\
&[T_2,A_\mu]=0,\quad [T_2,B_\mu]=0,\\
&[T_2,C_\mu]=iD_\mu,\quad [T_2,D_\mu]=-iC_\mu,\\
&[T_2,E_{\mu\nu}]=(i/2)\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta},\quad [T_2,F_{\mu\nu}]=(i/2)\epsilon_{\mu\nu\alpha\beta}E^{\alpha\beta},\\
&[E_{\mu\nu},E_{\alpha\beta}]=i(g_{\mu\alpha}S_{\nu\beta}-g_{\nu\alpha}S_{\mu\beta}+g_{\nu\beta}S_{\mu\alpha}-g_{\mu\beta}S_{\nu\alpha}),\qquad (3.2)\\
&[F_{\mu\nu},F_{\alpha\beta}]=-i(g_{\mu\alpha}S_{\nu\beta}-g_{\nu\alpha}S_{\mu\beta}+g_{\nu\beta}S_{\mu\alpha}-g_{\mu\beta}S_{\nu\alpha}),\\
&[E_{\mu\nu},F_{\alpha\beta}]=-i(g_{\mu\alpha}g_{\nu\beta}T_1-g_{\nu\alpha}g_{\mu\beta}T_1+\epsilon_{\mu\nu\alpha\beta}T_2),\\
&[A_\mu,A_\nu]=iS_{\mu\nu},\quad [A_\mu,B_\nu]=ig_{\mu\nu}T_1,\\
&[A_\mu,C_\nu]=iE_{\mu\nu},\quad [A_\mu,D_\nu]=(i/2)\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta},\\
&[B_\mu,B_\nu]=-iS_{\mu\nu},\quad [B_\mu,C_\nu]=-iF_{\mu\nu},\\
&[B_\mu,D_\nu]=-(i/2)\epsilon_{\mu\nu\alpha\beta}E^{\alpha\beta},\quad [C_\mu,C_\nu]=iS_{\mu\nu},\\
&[C_\mu,D_\nu]=ig_{\mu\nu}T_2,\quad [D_\mu,D_\nu]=iS_{\mu\nu},\\
&[A_\mu,E_{\alpha\beta}]=i(g_{\mu\beta}C_\alpha-g_{\mu\alpha}C_\beta),\quad [A_\mu,F_{\alpha\beta}]=i\epsilon_{\mu\alpha\beta\nu}D^\nu,\\
&[B_\mu,E_{\alpha\beta}]=i\epsilon_{\mu\alpha\beta\nu}D^\nu,\quad [B_\mu,F_{\alpha\beta}]=i(g_{\mu\beta}C_\alpha-g_{\mu\alpha}C_\beta),\\
&[C_\mu,E_{\alpha\beta}]=i(g_{\mu\beta}A_\alpha-g_{\mu\alpha}A_\beta),\\
&[C_\mu,F_{\alpha\beta}]=-i(g_{\mu\beta}B_\alpha-g_{\mu\alpha}B_\beta),\\
&[D_\mu,E_{\alpha\beta}]=-i\epsilon_{\mu\alpha\beta\nu}B^\nu,\quad [D_\mu,F_{\alpha\beta}]=i\epsilon_{\mu\alpha\beta\nu}A^\nu.
\end{aligned}
$$

The form of the Lie algebra has, of course, no dependence upon the particular identifications (3.1) used to obtain it, and any representation of the generators, e.g., by matrices, may be used in connection with (3.2). In the following section, the representational identities associated with the identifications (3.1) and Table II will be obtained.

4. OPERATOR IDENTITIES FOR THE BOSON REPRESENTATION

The ten generators of $\mathrm{Sp}(4,R)\equiv\mathrm{SO}(3,2)$ may be given a concrete bosonic representation in terms of q_1, q_2, η_1, and η_2, via Table II and

$$
\begin{aligned}
&V_\mu\equiv\tfrac{1}{2}(A_\mu+C_\mu),\\
&\Sigma_{\mu\nu}\equiv\tfrac{1}{2}(S_{\mu\nu}+E_{\mu\nu}).
\end{aligned}
\qquad (4.1)
$$

The operators $\Sigma_{\mu\nu}$ then generate the Majorana representation of the Lorentz group. It has been shown, in Ref. 9, that the following identities hold for this realization:

$$
\begin{aligned}
&V^\mu V_\mu=-\tfrac{1}{2},\\
&\Sigma^{\mu\nu}\Sigma_{\mu\nu}=-\tfrac{3}{2},\\
&\epsilon_{\mu\nu\alpha\beta}\Sigma^{\mu\nu}\Sigma^{\alpha\beta}=0.
\end{aligned}
\qquad (4.2)
$$

Now an identically similar representation of $\mathrm{Sp}(4,R)$ can be constructed in terms of q_3, q_4, η_3, and η_4 via Table II and

$$
\begin{aligned}
&\hat{V}_\mu\equiv\tfrac{1}{2}(A_\mu-C_\mu),\\
&\hat{\Sigma}_{\mu\nu}\equiv\tfrac{1}{2}(S_{\mu\nu}-E_{\mu\nu}).
\end{aligned}
\qquad (4.3)
$$

Therefore, the identities (4.2) apply as well to this set of operators, which commute with the first set.

When the identities (4.2) are applied in turn to the operators (4.1) and (4.3), the following preliminary identities result:

$$
\begin{aligned}
&A^\mu A_\mu+C^\mu C_\mu=-2,\\
&A^\mu C_\mu+C_\mu A^\mu=0,\\
&S^{\mu\nu}S_{\mu\nu}+E^{\mu\nu}E_{\mu\nu}=-6,
\end{aligned}
\qquad (4.4)
$$

$$S^{\mu\nu}E_{\mu\nu} + E_{\mu\nu}S^{\mu\nu} = 0,$$

$$\epsilon_{\mu\nu\alpha\beta}(S^{\mu\nu}S^{\alpha\beta} + E^{\mu\nu}E^{\alpha\beta}) = 0,$$

and

$$\epsilon_{\mu\nu\alpha\beta}(S^{\mu\nu}E^{\alpha\beta} + E^{\mu\nu}S^{\alpha\beta}) = 0.$$

The second of identities (4.4) taken together with the result, from (3.2),

$$[A^{\mu}, C_{\mu}] = 0, \tag{4.5}$$

yields immediately the result

$$A^{\mu}C_{\mu} = 0. \tag{4.6}$$

Successive commutations of (4.6) with T_1 and T_2 yields then

$$B^{\mu}C_{\mu} = 0,$$

$$A^{\mu}D_{\mu} = 0, \tag{4.7}$$

and

$$B^{\mu}D_{\mu} = 0.$$

Similarly, commutation of the first identity of (4.4) with T_1 and also T_2, and comparison of the results obtained with those of (3.2) yields the identities

$$A^{\mu}B_{\mu} = 2iT_1 \tag{4.8}$$

and

$$C^{\mu}D_{\mu} = 2iT_2.$$

Commutation of the first of identities (4.8) with A^{ν}, followed by a contraction of the result with B_{ν} and a second use of (3.2) yields the preliminary identity

$$S^{\mu\nu}S_{\mu\nu} + 4T_1^2 + 2A^{\mu}A_{\mu} + 4B^{\mu}B_{\mu} = 0, \tag{4.9a}$$

while the same procedure in the opposite order yields

$$S^{\mu\nu}S_{\mu\nu} + 4T_1^2 - 2A^{\mu}A_{\mu} = 0. \tag{4.9b}$$

Similarly, commutation of the second of identities (4.8) with C^{ν}, followed by a contraction of the result with D_{μ}, etc., and the same procedure in the opposite order yields the two results

$$S^{\mu\nu}S_{\mu\nu} - 4T_2^2 + 2C^{\mu}C_{\mu} - 4D^{\mu}D_{\mu} = 0 \tag{4.9c}$$

and

$$S^{\mu\nu}S_{\mu\nu} - 4T_2^2 - 2C^{\mu}C_{\mu} = 0. \tag{4.9d}$$

Comparison of the results (4.9), and use of the third of identities (4.4) then yields, finally,

$$\begin{aligned} A^{\mu}A_{\mu} &= -1 + T_1^2 + T_2^2, \\ B^{\mu}B_{\mu} &= 1 - T_1^2 - T_2^2, \\ C^{\mu}C_{\mu} &= -1 - T_1^2 - T_2^2, \\ D^{\mu}D_{\mu} &= -1 - T_1^2 - T_2^2, \\ S^{\mu\nu}S_{\mu\nu} &= -2 - 2T_1^2 + 2T_2^2, \end{aligned} \tag{4.10}$$

and

$$E^{\mu\nu}E_{\mu\nu} = -4 + 2T_1^2 - 2T_2^2.$$

The complete set of useful operator identities for the boson realization of $\mathcal{L}\,\mathrm{Sp}(8,R)$ is given in Table III. The procedure for obtaining any one of them is similar to that outlined above, and is left as an exercise for the interested reader.

These identities are essential for the purpose of constructing relativistic wave equations based on $\mathrm{Sp}(8,R)$ with particular physical content, after the fashion of Ref. 8. Subsidiary conditions, such as usually arise

TABLE III. Operator identities holding for the realization of the Lie algebra of $\mathrm{Sp}(8,R)$ by bosons employed.

$A^{\mu}A_{\mu} = -1 + (T_1^2 + T_2^2)$	$A^{\mu}B_{\mu} = 2iT_1$	$A^{\mu}C_{\mu} = 0$	$A^{\mu}D_{\mu} = 0$
$B^{\mu}B_{\mu} = 1 - (T_1^2 + T_2^2)$	$B^{\mu}A_{\mu} = -2iT_1$	$B^{\mu}C_{\mu} = 0$	$B^{\mu}D_{\mu} = 0$
$C^{\mu}C_{\mu} = -1 - (T_1^2 + T_2^2)$	$C^{\mu}A_{\mu} = 0$	$C^{\mu}B_{\mu} = 0$	$C^{\mu}D_{\mu} = 2iT_2$
$D^{\mu}D_{\mu} = -1 - (T_1^2 + T_2^2)$	$D^{\mu}A_{\mu} = 0$	$D^{\mu}B_{\mu} = 0$	$D^{\mu}C_{\mu} = -2iT_2$

$S^{\mu\nu}S_{\mu\nu} = -2 - 2T_1^2 + 2T_2^2$	$S^{\mu\nu}E_{\mu\nu} = 0$	$S^{\mu\nu}F_{\mu\nu} = 0$
$E^{\mu\nu}E_{\mu\nu} = -4 + 2T_1^2 - 2T_2^2$	$E^{\mu\nu}S_{\mu\nu} = 0$	$E^{\mu\nu}F_{\mu\nu} = -6iT_1$
$F^{\mu\nu}F_{\mu\nu} = 4 - 2T_1^2 + 2T_2^2$	$F^{\mu\nu}S_{\mu\nu} = 0$	$F^{\mu\nu}E_{\mu\nu} = 6iT_1$

$\epsilon_{\mu\nu\alpha\beta}S^{\mu\nu}S^{\alpha\beta} = 8T_1T_2$	$\epsilon_{\mu\nu\alpha\beta}S^{\mu\nu}E^{\alpha\beta} = 0$	$\epsilon_{\mu\nu\alpha\beta}S^{\mu\nu}F^{\alpha\beta} = 0$
$\epsilon_{\mu\nu\alpha\beta}E^{\mu\nu}E^{\alpha\beta} = -8T_1T_2$	$\epsilon_{\mu\nu\alpha\beta}E^{\mu\nu}S^{\alpha\beta} = 0$	$\epsilon_{\mu\nu\alpha\beta}E^{\mu\nu}F^{\alpha\beta} = 12iT_2$
$\epsilon_{\mu\nu\alpha\beta}F^{\mu\nu}F^{\alpha\beta} = 8T_1T_2$	$\epsilon_{\mu\nu\alpha\beta}F^{\mu\nu}S^{\alpha\beta} = 0$	$\epsilon_{\mu\nu\alpha\beta}F^{\mu\nu}E^{\alpha\beta} = -12iT_2$

$A^{\mu}S_{\mu\nu} = B_{\nu}T_1 - iA_{\nu}$	$A^{\mu}E_{\mu\nu} = D_{\nu}T_2 - 2iC_{\nu}$	$A^{\mu}F_{\mu\nu} = C_{\nu}T_1$
$B^{\mu}S_{\mu\nu} = A_{\nu}T_1 - iB_{\nu}$	$B^{\mu}E_{\mu\nu} = C_{\nu}T_1$	$B^{\mu}F_{\mu\nu} = D_{\nu}T_2 - 2iC_{\nu}$
$C^{\mu}S_{\mu\nu} = -D_{\nu}T_2 - iC_{\nu}$	$C^{\mu}E_{\mu\nu} = -B_{\nu}T_1 - 2iA_{\nu}$	$C^{\mu}F_{\mu\nu} = A_{\nu}T_1 + 2iB_{\nu}$
$D^{\mu}S_{\mu\nu} = C_{\nu}T_2 - iD_{\nu}$	$D^{\mu}E_{\mu\nu} = A_{\nu}T_2$	$D^{\mu}F_{\mu\nu} = -B_{\nu}T_2$

$\epsilon_{\mu\nu\alpha\beta}A^{\nu}S^{\alpha\beta} = 2B_{\mu}T_2$	$\epsilon_{\mu\nu\alpha\beta}A^{\nu}E^{\alpha\beta} = -2D_{\mu}T_1$	$\epsilon_{\mu\nu\alpha\beta}A^{\nu}F^{\alpha\beta} = 2C_{\mu}T_2 + 4iD_{\mu}$
$\epsilon_{\mu\nu\alpha\beta}B^{\nu}S^{\alpha\beta} = 2A_{\mu}T_2$	$\epsilon_{\mu\nu\alpha\beta}B^{\nu}E^{\alpha\beta} = -2C_{\mu}T_2 - 4iD_{\mu}$	$\epsilon_{\mu\nu\alpha\beta}B^{\nu}F^{\alpha\beta} = -2D_{\mu}T_1$
$\epsilon_{\mu\nu\alpha\beta}C^{\nu}S^{\alpha\beta} = 2D_{\mu}T_1$	$\epsilon_{\mu\nu\alpha\beta}C^{\nu}E^{\alpha\beta} = -2B_{\mu}T_2$	$\epsilon_{\mu\nu\alpha\beta}C^{\nu}F^{\alpha\beta} = 2A_{\mu}T_2$
$\epsilon_{\mu\nu\alpha\beta}D^{\nu}S^{\alpha\beta} = -2C_{\mu}T_1$	$\epsilon_{\mu\nu\alpha\beta}D^{\nu}E^{\alpha\beta} = -2A_{\mu}T_1 - 4iB_{\mu}$	$\epsilon_{\mu\nu\alpha\beta}D^{\nu}F^{\alpha\beta} = 2B_{\mu}T_1 + 4iA_{\mu}$

$$A_{\mu}A_{\nu} - B_{\mu}B_{\nu} - C_{\mu}C_{\nu} - D_{\mu}D_{\nu} = g_{\mu\nu}(T_1^2 + T_2^2)$$

when constrained Hamiltonian systems are considered, may be formulated in terms of the Lorentz scalar operators T_1 and T_2 in this way.

For example, the operator A_0, from Table II, has the number operator form

$$A_0 = \tfrac{1}{2}(\bar{a}_1 a_1 + \bar{a}_2 a_2 + \bar{a}_3 a_3 + \bar{a}_4 a_4) + 1, \tag{4.11}$$

where the operators a and $\bar{a}$ denote boson annihilation and creation operators, respectively, while the operator T_2 has the form

$$T_2 = (i/2)(\bar{a}_3 a_1 - \bar{a}_1 a_3 + \bar{a}_4 a_2 - \bar{a}_2 a_4). \tag{4.12}$$

If the subsidiary condition on the eigenstates of A_0 is imposed that T_2 must annihilate physical states, then the physical states are restricted to those containing only even numbers of quanta, i.e., a signature factor effect is automatically incorporated.

Details of applications will be the subject of another communication.

ACKNOWLEDGMENTS

The author is grateful for beneficial conversations with Professor Granville Smith of Grinnel College, and to the staff of the Dial Computer Center of Drake University for assistance with algebraic manipulation codes.

[1]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).

[2]P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963); G. Loupias, M. Sirugue and J. C. Trotin, Nuovo Cimento **38**, 1303 (1965).

[3]E. Majorana, Nuovo Cimento **9**, 335 (1932). An account in English is given by D. M. Fradkin, Am. J. Phys. **34**, 314 (1966). See also the review lecture of A. Böhm, in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. X-B.

[4]See, for example, P. H. Frampton, *Dual Resonance Models* (Benjamin, New York, 1974).

[5]Y. Nambu, Lectures at the Copenhagen Summer Symposium, 1970 (unpublished).

[6]P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, Nucl. Phys. B **56**, 109 (1973).

[7]R. Marnelius, Nucl. Phys. B **104**, 477 (1976).

[8]L. P. Staunton, Phys. Rev. D **10**, 1760 (1974).

[9]L. P. Staunton and S. Browne, Phys. Rev. D **12**, 1026 (1975).

[10]L. C. Biedenharn, M. Y. Han, and H. van Dam. Phys. Rev. D **8**, 1735 (1973).

[11]L. P. Staunton, Phys. Rev. D **13**, 3269 (1976).

[12]The beginnings of a complete string model from a somewhat related but alternative, postulatory approach have been reported by F. Rohrlich, Phys. Rev. D **16**, 354 (1977), and earlier references therein.

[13]S. Browne, Nucl. Phys. B **79**, 70 (1974). The result is not restricted to real matrices, nor to ordinary (nongraded) Lie algebras.

[14]Note carefully that these results are for the operators G defined in (2.12). The commutation relations among the matrices Γ themselves are different from those of (2.13).

[15]P. A. M. Dirac, Proc. R. Soc. London A **322**, 435 (1971); **328**, 1 (1972).

[16]The operators (3.1), as well as those others listed in Table II are self-adjoint under the inner product $\int d^4q$.

[17]L. P. Staunton, Phys. Rev. D **8**, 2446 (1973); see also Ref. 10.

[18]Our conventions are $\hbar = c = 1$, $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$, and $\epsilon_{0123} = +1$.

The algebra of color[a)]

G. Domokos and S. Kövesi-Domokos

Department of Physics, The Johns Hopkins University, Baltimore, Maryland 21218
(Received 9 February 1978)

We construct an algebra of (local) dynamical variables which satisfies the triality rule for quarks and an exact superselection rule between leptons and quarks. This algebra is isomorphic to a noncommutative Jordan algebra; hence, it is power associative. Inner derivations are constructed explicitly: Their algebra is isomorphic to Lie SU(3); the latter can be identified with the group of color symmetry.

1. INTRODUCTION

Color is a peculiar symmetry in several respects. First, it appears that all observables are singlets under the color group, $SU(3)_c$, even though some dynamical variables (quarks) transform according to nontrivial representations of the color group.[1] Second, it seems that color is an *exact* local symmetry. Third, only a very limited number of the possible representations of $SU(3)_c$ seems to play a role in the description of hadrons. For this reason, it has been proposed that in any unified theory involving leptons and quarks as the fundamental fermionic variables, the color properties should be reflected in the algebraic rules which govern the multiplication of the dynamical variables.[2] [In this respect, $SU(3)_c$ would play a role analogous to the role played by the symmetric group in the formulation of the Pauli principle.]

To be specific, we assume, as usual, that quarks are triplets (3), antiquarks are antitriplets ($\bar{3}$), and leptons are singlets (1) under $SU(3)_c$. The multiplication rule of the corresponding dynamical variables should be such that "mesons" (quark–antiquark) and "baryons" (three quarks) are both singlets under $SU(3)_c$. This can be achieved[2] if the multiplication table of the dynamical variables obeys the *triality rule*, which symbolically can be written as follows:

$$(3)\times(\bar{3})\sim(1),\quad (3)\times(3)\sim(\bar{3}). \tag{1.1}$$

Notice that (1.1) merely states the *absence* of certain representation of $SU(3)_c$ from the product of two dynamical variables: The triality rule says nothing about the products of, say, quarks and leptons. Gürsey and his collaborators found realizations of (1.1) in terms of split-octonion valued variables,[3] once this choice is made, the rest of the multiplication rules follow.

The implementation of the triality rule by means of octonion-valued variables is, however, not the only possibility, nor is it necessarily the most desirable one from the physical point of view.[4] Instead of committing ourselves to any predetermined algebra, we argue that the algebra of the dynamical variables should be constructed by means of a step-by-step implementation of physical principles. The triality rule, (1.1) is among the principles we want to incorporate, since it appears to be dictated by particle phenomenology. However, additional physical input is needed in order to complete the algebra. In particular, we construct the algebra in a way which incorporates an *exact superselection rule between hadrons and leptons*.

The paper is organized as follows. In Sec. 2 we work out a simple exercise: para-Fermi fields realized on a direct product algebra. This exercise is of no particular physical significance, however, it gives a clue to the form of dynamcial variables on which we want to implement the physical principles. The actual color algebra is constructed in Sec. 3, whereas its structure, including the algebra of inner derivations,[5] is described in Sec. 4. In Sec. 5 we sketch an interesting construction due to Faulkner. His procedure enables one to build up a large number of "color algebras," among those the color algebra proposed in this work appears as a (perhaps, physically important) special case. Finally, in Sec. 6 we endow our color algebra with a grading. The existence of a grading is physically plausible: Throughout this work we regard dynamical variables describing quarks and leptons as generalized Fermi variables.

In this work we concentrate on the correct algebraic description of color. Other symmetries (flavor and space–time groups) are not explicitly touched upon, even though their presence is implicitly assumed. In this sense, the present work may be regarded as a first step towards the construction of a physically viable unified theory of fundamental interactions: Subsequent steps obviously have to include a unification of color and flavor degrees of freedom and the construction of appropriate bundles in terms of local variables obeying physically relevant algebraic relations.

2. EXCEPTIONAL REALIZATION OF PARA-FERMI VARIABLES: A PEDAGOGICAL EXERCISE

Consider a para-Fermi field, ψ, of order p, realized through the Green ansatz.[6] On suppressing space–time variables and possible internal symmetry indices, we may write

$$\psi=\sum_{\alpha=1}^{p}\psi^{(\alpha)}, \tag{2.1}$$

where the Green components, $\psi^{(\alpha)}$, obey the following well-known algebraic relations:

$$\begin{aligned}&\psi^{(\alpha)}\psi^{(\beta)}+(-1)^{1+\delta_{\alpha,\beta}}\psi^{(\beta)}\psi^{(\alpha)}=0,\\ &\psi^{(\alpha)\dagger}\psi^{(\beta)}+(-1)^{1+\delta_{\alpha,\beta}}\psi^{(\beta)}\psi^{(\alpha)\dagger}=\delta_{\alpha,\beta}E,\end{aligned} \tag{2.2}$$

where E stands for the unit operator and Hermitian conjugates are denoted by a dagger.

[a)]Research supported in part by the U. S. Department of Energy under contract No. EY-76-S-02-3285.

 0022-2488/78/1906-1477$1.00

Conventionally, the Green components are written in terms of ordinary Fermi fields, $\phi^{(\alpha)}$, via a Klein transformation[7]:

$$\psi^{(\alpha)} = (-i)^{1-\rho(\alpha)} K_{\alpha+\rho(\alpha)} \phi^{(\alpha)} \tag{2.3}$$

with

$$\rho(\alpha) \equiv \tfrac{1}{2}[1+(-1)^{\alpha+1}].$$

Here, as usual,

$$\{\phi^{(\alpha)}, \phi^{(\beta)}\} = 0, \quad \{\phi^{(\alpha)\dagger}, \phi^{(\beta)}\} = \delta_{\alpha,\beta} E, \tag{2.4}$$

and

$$\begin{aligned} &[K_\beta, \phi^{(\gamma)}] = 0 \quad (\beta > \gamma), \\ &\{K_\beta, \phi^{(\gamma)}\} = 0 \quad (\beta \leq \gamma), \\ &K_\beta = K_\beta^\dagger = K_\beta^{-1}, \end{aligned} \tag{2.5}$$

cf. e.g., Drühl *et al.*[8] The important point about this realization is that the Klein operators, K_α, are functionals of the Fermi variables $\phi^{(\alpha)}$ and their Hermitian conjugates.

The latter assumption may be relaxed; however, in this way one arrives at inequivalent realizations of the relations (2.2).

Indeed, consider the following ansatz for a para-Fermi field (of order p),

$$\psi = \sum_{\alpha=1}^{p} e_\alpha \phi^{(\alpha)} \equiv \sum_{\alpha=1}^{p} \psi^{(\alpha)}, \tag{2.1'}$$

where the $\phi^{(\alpha)}$ are Fermi variables obeying (2.4). However, instead of (2.5) we require

$$[e_\alpha, \phi^{(\beta)}] = 0 \qquad \forall\, \alpha, \beta. \tag{2.6}$$

It is an elementary excercise to show that the Green relations, (2.2) are satisfied if the e_α generate a (real) Clifford algebra, viz.,

$$\{e_\alpha, e_\beta\} = \delta_{\alpha\beta}. \tag{2.7}$$

We realized para-Fermi variables on a *direct product algebra* such that the direct factors form a real Clifford algebra $C(p,R)$ and an ordinary Fermi algebra, (2.4), respectively.

By inspecting (2.1'), we discover that ψ possesses a "color group,[9]" which is isomorphic to Aut $C(p,R) \approx \mathrm{SO}(p,R)$. Given the fact that $\mathrm{SO}(p,R)$ is a *real* group, for the sake of consistency we have to impose a Majorana condition on the $\phi^{(\alpha)}$, viz.,

$$C\phi^{(\alpha)} = \phi^{(\alpha)},$$

where C stands for the charge conjugation operators. We notice the important fact that the "color" group is generated by the algebra of inner derivations of $C(p, R)$. Explicitly, if $O_{\alpha\beta}$ stands for a conventional basis of Lie $\mathrm{SO}(p,R)$, such that

$$\begin{aligned} [O_{\alpha\beta}, O_{\gamma\delta}] = {} & \delta_{\alpha\delta} O_{\beta\gamma} + \delta_{\beta\gamma} O_{\alpha\delta} \\ & - \delta_{\alpha\gamma} O_{\beta\delta} - \delta_{\rho\delta} O_{\alpha\gamma}, \end{aligned} \tag{2.8}$$

then $O_{\alpha\beta}$ is realized by

$$O_{\alpha\beta} : e_\gamma \to \tfrac{1}{2}[[e_\alpha, e_\beta], e_\gamma]. \tag{2.9}$$

In a sense, this exercise is a trivial one: It is unlikely that Majorana para-Fermi fields are of any physical interest. The important lesson to be learnt, however, is that color groups can be realized as inner automorphisms on dynamical variables constructed on the direct product of appropriate algebras.

3. CONSTRUCTION OF THE COLOR ALGEBRA

Motivated by the construction of para-Fermi fields in terms of the ansatz (2.1'), (2.6), we now want to construct dynamical variables which describe the color properties of quarks and leptons in a unified way. To this end, we make an ansatz for the fundamental dynamical variable ψ, analogous to (2.1'), viz.,

$$\psi = u_0 l + u_\alpha q^\alpha \tag{3.1}$$

with $1 \leq \alpha \leq 3$; summation over repeated Greek indices ("color indices") is understood. The coefficients l and q^α are to represent leptonic and quark variables, respectively: They are regarded as ordinary, anticommuting, Fermi variables. Space—time and internal symmetry (flavor) labels are suppressed as before. Hence all the color properties and the triality rule in particular, have to be realized in the basis elements u_0 and u_α. We now proceed to implement some elementary physical requirements in order to determine the multiplication table of these basis elements.

(i) Quarks and antiquarks (leptons and antileptons) are distinct. Both leptons and antileptons transform as $\sim(1)$ under $\mathrm{SU}(3)_c$. However, quarks transform as $\sim(3)$, whereas antiquarks transform as $\sim(\bar{3})$. Hence, we must have a dinstinct set of basis elements, $\bar{u}_\alpha$, such that as basis vectors of a vector space, u_α span (3) and $\bar{u}_\alpha$ span $(\bar{3})$ of $\mathrm{SU}(3)_c$. We write the *conjugate* (for instance, under TCP) of ψ as

$$\bar{\psi} = \bar{u}_0 \bar{l} + \bar{u}_\alpha \bar{q}_\alpha, \tag{3.2}$$

where both u_0 and $\bar{u}_0$ are singlets under $\mathrm{SU}(3)_c$.

(ii) Hadron-lepton superselection rule: Every observable, Ω, may be written as a direct sum with the help of orthogonal projectors (idempotents), viz.,

$$\Omega = L\Omega L + H\Omega H \tag{3.3}$$

such that

$$\begin{aligned} &L^2 = L, \quad H^2 = H, \\ &L^\dagger = L, \quad H^\dagger = H, \\ &LH = HL = 0, \end{aligned} \tag{3.4}$$

and, naturally, both L and H are singlets under $\mathrm{SU}(3)_c$. Since we want to identify l with a leptonic variable (an observable), u_0 may be identified with the projector L, provided we postulate $\bar{u}_0 = u_0$; this is obviously permissible, since both u_0 and $\bar{u}_0$ are color singlets. We also assume that there are just two superselection sectors (hadrons and leptons), so that

$$L + H = E, \tag{3.5}$$

where E is the unit element of the color algebra. The last relation allows us to write $L = E - H$, so that we do not have to consider L and H separately.

Furthermore, leptons and quarks fall into different superselection sectors; hence, we must have (with

$u=\bar{u}_0=L$),

$$Lu_\alpha=u_\alpha L=L\bar{u}_\alpha=\bar{u}_\alpha L=0. \tag{3.6}$$

(iii) Triality rule: Mesons (or currrents) must be observables, and hence, of the form (3.3). Keeping in mind (3.1), (3.2), (3.6), we thus have

$$u_\alpha\bar{u}_\beta=\bar{u}_\alpha u_\beta=\delta_{\alpha\beta}H. \tag{3.7}$$

[In writing down (3.7), we kept in mind flavor and space–time properties to be carried by the Fermi coefficients q_α and $\bar{q}_\alpha$: These must be the same as predicted by a "naive" quark model.] Similarly, baryons (antibaryons) must be hadronic observables. Together with (3.7), this can be achieved by putting

$$u_\alpha u_\beta=\epsilon_{\alpha\beta\gamma}\bar{u}_\gamma,\quad \bar{u}_\alpha\bar{u}_\beta=\epsilon_{\alpha\beta\gamma}u_\gamma, \tag{3.8}$$

where $\epsilon_{\alpha\beta\gamma}$ is the totally antisymmetric unit tensor.

(iv) Finally, any hadronic observable and a quark (antiquark) must carry the same color as the quarks do, with the correct flavor and space–time properties. This gives

$$Hu_\alpha=u_\alpha H=u_\alpha,\quad H\bar{u}_\alpha=\bar{u}_\alpha H=\bar{u}_\alpha. \tag{3.9}$$

This completes the multiplication table of the basis spanned by u_α, $\bar{u}_\alpha$, and H. [Obviously, the multiplication rules of L with the other basic elements can be read off with the help of (3.5).] To summarize, we have the multiplication table:

	H	u_β	$\bar{u}_\beta$
H	H	u_β	$\bar{u}_\beta$
u_α	$\bar{u}_\alpha$	$\epsilon_{\alpha\beta\gamma}\bar{u}_\gamma$	$\delta_{\alpha\beta}H$
$\bar{u}_\alpha$	$\bar{u}_\alpha$	$\delta_{\alpha\beta}H$	$\epsilon_{\alpha\beta\gamma}u_\gamma$

(3.10)

We observe that the multiplication table (3.10) is unique, up to linear transformations of the basis elements.

4. STRUCTURE OF THE ALGEBRA AND DERIVATIONS

The algebra defined by the multiplication table (3.10) is a simple nonassociative algebra; henceforth it will be referred to as $\mathcal{S}$.

The table of nonvanishing associators can be worked out with the help of (3.10). We find $\{[a,b,c]\equiv(ab)c-a(bc)\}$:

$$\begin{aligned}
&[\bar{u}_\beta,u_\gamma,u_\alpha]=-[u_\alpha,u_\gamma,\bar{u}_\beta]=2\delta_{\beta\gamma}u_\alpha-\delta_{\alpha\beta}u_\gamma,\\
&[u_\gamma,u_\alpha,\bar{u}_\beta]=-[\bar{u}_\beta,u_\alpha,u_\gamma]=\delta_{\beta\gamma}u_\alpha-2\delta_{\alpha\beta}u_\gamma,\\
&[u_\alpha,\bar{u}_\beta,u_\gamma]=-[u_\gamma,\bar{u}_\beta,u_\alpha]=\delta_{\alpha\beta}u_\gamma-\delta_{\beta\gamma}u_\alpha,\\
&[\bar{u}_\beta,u_\gamma,\bar{u}_\alpha]=-[\bar{u}_\alpha,u_\gamma,\bar{u}_\beta]=\delta_{\beta\gamma}\bar{u}_\alpha-\delta_{\gamma\alpha}\bar{u}_\beta,\\
&[u_\gamma,\bar{u}_\alpha,\bar{u}_\beta]=-[\bar{u}_\beta,\bar{u}_\alpha,u_\gamma]=2\delta_{\alpha\gamma}\bar{u}_\beta-\delta_{\beta\gamma}\bar{u}_\alpha,\\
&[\bar{u}_\alpha,\bar{u}_\beta,u_\gamma]=-[u_\gamma,\bar{u}_\beta,\bar{u}_\alpha]=\delta_{\alpha\beta}\bar{u}_\beta-2\delta_{\beta\gamma}\bar{u}_\alpha.
\end{aligned} \tag{4.1}$$

We immediately verify that $\mathcal{S}$ is not an alternative algebra, since for a generic pair of x, $y\in\mathcal{S}$, $[x,x,y]\neq 0$. Nevertheless, $\mathcal{S}$ is *flexible*, i.e., $[x,y,x]\equiv 0$, as one can verify by direct computation.

Form a physical point of view it is of utmost importance to ascertain that $\mathcal{S}$ is *power associative*, i.e., the subalgebra generated by any fixed $x\in\mathcal{S}$ is associative. This property allows us to use the variables ψ and $\bar{\psi}$ as dynamical variables; in particular polynomials of the dynamical variables are defined unambiguously. In order to verify power associativity, we resort to the well-known tool[10] of passing from $\mathcal{S}$ to its symmetrized version, $\mathcal{S}^+$. The algebra $\mathcal{S}^+$ is defined to be the same vector space as $\mathcal{S}$, but with product defined by

$$x\circ y\equiv\tfrac{1}{2}(xy+yx).$$

The multiplication table of $\mathcal{S}^+$ can be read off from (3.10):

$\mathcal{S}^+$:

	H	u_β	$\bar{u}_\beta$
H	H	u_β	$\bar{u}_\beta$
u_α	u_α	0	$\delta_{\alpha\beta}H$
$\bar{u}_\alpha$	$\bar{u}_\alpha$	$\delta_{\alpha\beta}H$	0

(3.10+)

We also remark in passing that the antisymmetrized algebra $\mathcal{S}^-$, with product defined by $x\wedge y\equiv\frac{1}{2}(xy-yx)$, is just a Mal'cev algebra with identity, as it can be readily seen from (3.10) after antisymmetrization:

$\mathcal{S}^-$:

	H	u_β	$\bar{u}_\beta$
H	0	0	0
u_α	0	$\epsilon_{\alpha\beta\gamma}\bar{u}_\gamma$	0
u_α	0	0	$\epsilon_{\alpha\beta\gamma}u_\gamma$

(3.10−)

Returning to $\mathcal{S}^+$, we immediately verify that $\mathcal{S}^+$ is a (commutative) Jordan algebra; for instance, one verifies the Jordan identities[11] by direct computation,

$$[x^2,y,x]=[x^2,x,y]=[y,x,x^2]=0,\quad \forall\ x,y\in\mathcal{S}^+. \tag{4.2}$$

It folows then that $\mathcal{S}$ is a non-commutative Jordan algebra[12] and, hence, it is power associative.

Automorphisms and involutions: The following statements are verified by inspection of (3.10).

(a) Discrete transformations:

$$u_\alpha\longleftrightarrow\bar{u}_\alpha,\quad H\to H \tag{4.3}$$

is an automorphism, whereas

$$u_\alpha\longleftrightarrow-\bar{u}_\alpha,\quad H\longleftrightarrow H \tag{4.4}$$

is an involution, i.e., for $x,y\in\mathcal{S}$, $(\overline{xy})=\bar{y}\bar{x}$ under (4.4). The existence of these discrete transformations allows the construction of two nontrivial reflections (for instance, C and T) on the variables ψ and $\bar{\psi}$.

(b) Continuous automorphisms: The linear transformations,

$$u'_\alpha=M_{\alpha\beta}u_\beta,\quad \bar{u}'_\alpha=M^*_{\alpha\beta}\bar{u}_\beta,\quad H'=H \tag{4.5}$$

with $M_{\alpha\beta}M^*_{\gamma\beta}=\delta_{\alpha\gamma}$, is an automorphism of (3.10), where M is a 3×3 complex unimodular, unitary matrix, the complex conjugate of a number being denoted by an asterisk. This is the built-in automorphism group of the algebra, which we identify with the color group, $SU(3)_c$.

It can be shown that SU(3) is actually the *largest* group of continuous automorphisms of $\mathcal{S}$ which are connected

with the identity automorphism.[13] We postpone the proof of this statement to the next section, where it will be carried out in the framework of an elegant construction due to Faulkner,[14]

From a physical point of view, it is important to see whether the infinitesimal transformations of $SU(3)_c$ can be realized as innner derivations on $\mathcal{S}$. (This is necessary in order to construct couplings between spinor and gauge fields in a theory where our variables ψ and $\bar{\psi}$ will be regarded as local ones, cf. Ref. 15). To this end, we introduce a basis, G^α_β in Lie SU(3), acting on u_α, $\bar{u}_\alpha$, and H in the usual way, viz.,

$$\begin{aligned} G^\beta_\gamma u_\alpha &= -(\delta^\beta_\alpha u_\gamma - \tfrac{1}{3}\delta^\beta_\gamma u_\alpha), \\ G^\beta_\gamma \bar{u}_\alpha &= \delta^\alpha_\gamma \bar{u}_\beta - \tfrac{1}{3}\delta^\beta_\gamma \bar{u}_\alpha, \\ G^\beta_\gamma H &= 0. \end{aligned} \tag{4.6}$$

It is now easily seen that the action of the G^β_γ is realized in terms of a linear combination of associators,

$$\begin{aligned} G^\beta_\gamma x \to \tfrac{1}{6}([x,u_\gamma,\bar{u}_\beta] + [u_\gamma,\bar{u}_\beta,x] \\ - 3[\bar{u}_\beta,x,u_\gamma]), \quad \forall\, x \in \mathcal{S}. \end{aligned} \tag{4.7}$$

One readily verifies that the operation (4.7) is indeed a derivation and that the Jacobi identities are satisfied as they should be.

5. FAULKNER'S CONSTRUCTION

Faulkner has constructed a class of noncommutative Jordan algebras; in this context, $\mathcal{S}$ appears as a special case belonging to that class.[14] Due to its intrinsic interest, we reproduce here the construction; in particular, this allows one to prove that the algebra of inner derivations of $\mathcal{S}$ is indeed isomorphic to the Lie algebra of SU(3). (This proof is also due to Faulkner; the proof given in Ref. 13 is essentially equivalent to the one reproduced here.)

First, a large class of noncommutative Jordan algebras may be constructed in the following way. We consider a vector space, V, with coefficients taken from an algebraic field, F. (For practical purposes, F may be the field of real or complex numbers.) We equip V with a nondegenerate symmetric scalar product, f, such that $f(x,y)=f(y,x)$ and $f(x,y)=0\ \forall x \in V$ implies $y=0$. Also, we define an alternating form, $g(u,v,w)$, on the outer product $V\wedge V\wedge V$. With the help of these, we may define an antisymmetric product on V, say $u\times v$ $(u,v\in V)$ such that $f(u\times u,w)=g(u,v,w)$.

{Let us remark at this point that such a construction is *not empty*. As an example, consider a simple Lie algebra with a basis a_i, normalized in such a way that the Killing form is $k(a_i,a_j)=\delta_{ij}$. Furthermore, let us take the Lie bracket to be of the form $[a_i,a_j]=C_{ijk}a_k$, where the structure constants C_{ijk} are totally antisymmetric. We can now identify $f\to k$, $a_i\times a_j \to [a_i a_j]$ and obviously, $g(a_i,a_j,a_k)=k([a_i,a_j],a_k)$.} We consider now a scalar extension of V, by adjoining a unit element, h, to it; as a vector space, we thus have $\mathcal{A}=Fh+V$. We promote $\mathcal{A}$ to an algebra by defining the product between elements of the form $(\alpha h+v)$ $(\alpha\in F,\ v\in V)$, as follows,

$$(\alpha h+v)(\beta h+u)=(\alpha\beta+f(v,u))h+(\alpha u+\beta v+v\times u). \tag{5.1}$$

By using the same tools as in the previous section, it is readily established that $\mathcal{A}$ is a noncommutative Jordan algebra with the product defined by (5.1). Indeed, the symmetrized algebra, $\mathcal{A}^+$ is easily shown to be a commutative Jordan algebra. [One has to verify the Jordan identities, (4.2), by direct computation.] Furthermore, by computing associators, the flexible law, $[a,b,a]=0$ can be verified.

Next, one inquires about derivations of $\mathcal{A}$. The basic point is to show that the derivations preserve the forms f and g.

The proof proceeds as follows. Let D be a derivation of $\mathcal{A}$; obviously, $D(h)=0$. Now, for $v\in V$, $v^2=f(v,v)h$, by (5.1). Hence, $D(v^2)=vD(v)=D(v)v=0$ and by taking scalar products, we have $f(v,D(v))=0$. In a similar way we find

$$f(D(u),v)+f(v,D(u))=0. \tag{5.2}$$

Thus D preserves scalar products. By an entirely similar reasoning, one verifies

$$D(v\times u)=D(v)\times u+v\times D(u).$$

From this, it follows that

$$g(D(u),v,w)+g(u,D(v),w)+g(u,v,D(w))=0. \tag{5.3}$$

Hence, derivations preserve f and g as stated.

In order to recover $\mathcal{S}$ as a special case, consider $V=U_3\oplus\bar{U}_3$ where U_3 and $\bar{U}_3$ are isomorphic copies of a three-dimensional vector space. Hence, for each $v\in V$, we may write $v=u_1+\bar{u}_2$, with $u_1\in U$, $u_2\in\bar{U}$. Furthermore, for each $u\in U$, there exists an image in $\bar{U}$ denoted by $\bar{u}$, and vice versa. We now take

$$\begin{aligned} f(u,\bar{v})=f(\bar{v},u)=u\cdot v h, \\ f(u,u)=f(\bar{u},\bar{u})=0, \end{aligned} \tag{5.4}$$

where $u\cdot v$ stands for the ordinary (Cartesian) scalar product. As an alternating form, we take

$$\begin{aligned} g(u,v,w)&=\det(u,v,w)h, \\ g(\bar{u},\bar{v},\bar{w})&=\det(u,v,w)h, \\ g(u,\bar{v},\bar{w})&=g(u,v,\bar{w})=0. \end{aligned} \tag{5.5}$$

Hence, the skew-symmetric product with $v=v_1+\bar{v}_2$, $w=w_1+\bar{w}_2$ $(v,w\in V)$ is

$$v\times w=\overline{v_1\times w_1}+v_2\times w_2, \tag{5.6}$$

where, on the rhs of (5.6) we have the ordinary cross product between Cartesian 3-vectors.

We remark further that our ansatz for V amounts to a complexification of a real, three-dimensional vector space; in order to see that, one may take, e.g., the vectors $u+\bar{u}$ and $u-\bar{u}$ from V.

Under the product (5.1), the multiplication table now becomes

	h	v	$\bar{v}$
h	h	v	$\bar{v}$
u	u	$\overline{u\times v}$	$(u\cdot v)h$
$\bar{u}$	$\bar{u}$	$(u\cdot v)h$	$u\times v$

(5.7)

which is obviously equivalent to (3.10). The derivation algebra preserves (5.4) and (5.5), hence it is isomorphic to Lie SU(3). Q.E.D.

Obviously, $\mathcal{S}$ fits into Faulkner's construction because the hadronic projector, H, acts as the unit element on the sector built up of quarks.

6. GRADING OF THE COLOR ALGEBRA

We claim that $\mathcal{S}$ can be given the structure of a graded algebra.[16] In order to verify this, we choose the basis used before, spanned by u_α, $\bar{u}_\alpha$, H. The grading group turns out to be Z_3. It is convenient to realize this group explicitly by taking the third roots of unity in the complex plane, viz.,

$$g_0 = 1, \quad g_1 = \exp(i2\pi/3), \quad g_2 = \exp(i4\pi/3). \tag{6.1}$$

In what follows, this ecplicit relaization is always implied when necessary. (In particular, it makes sense to *add* the elements of Z_3 as complex numbers.) We have the multiplication table:

	g_0	g_1	g_2
g_0	g_0	g_1	g_2
g_1	g_1	g_2	g_0
g_2	g_2	g_0	g_1

(6.2)

Correspondingly, we grade the vector space, $\mathcal{S} = h \oplus u \oplus \bar{u}$ as follows:

$$\text{grade}(h) = g_0, \quad \text{grade}(u) = g_1, \quad \text{grade}(\bar{u}) = g_2. \tag{6.3}$$

This map, $\mathcal{S} \to Z_3$, establishes an isomorphism between the multiplication tables (3.10) [or (5.7)] and (6.2). Hence, the *algebra* $\mathcal{S}$ is graded by Z_3 as asserted.

With the help of this grading, one can rewrite the multiplication table (3.10) in complete analogy with the multiplication table of the familiar Z_2 graded algebras. Indeed, let $u(g_i)$ stand for an element of a vector space ($\in V$) of grade g_i. We now have

$$\tfrac{1}{2}[u(g_i)u(g_j) + (-1)^{\sigma(g_i,g_j)}u(g_j)u(g_i)] = u(g_ig_j),$$
$$u(g_i)u(g_j) - (-1)^{\sigma(g_ig_j)}u(g_j)u(g_i) = 0, \tag{6.4}$$

where

$$\sigma(g_i, g_j) = -\tfrac{2}{3}\text{Re}(g_i + g_j + g_ig_j). \tag{6.5}$$

This completes the grading process.

ACKNOWLEDGMENTS

We wish to thank G. Benkart, J.R. Faulkner, F. Gürsey, J.M. Osborn, and V. Rittenberg for illuminating discussions and correspondence on the subject. We also thank G. Benkart, J.R. Faulkner and J.M. Osborn for letting us know about several of their results regarding the color algebra prior to publication. This work has been completed during the authors' visit at the Stanford Linear Accelerator Center. We thank J.D. Bjorken and S.D. Drell for the hospitality extended to us at SLAC.

[1]We assume, as usual, that the color group is isomorphic to SU(3). For other possibilities, cf. S. Okubo, Phys. Rev. D **16**, 3535 (1977).
[2]F. Gürsey, in The Proceedings of the Johns Hopkins University Workshop, edited by G. Domokos and S. Kövesi-Domokos, Baltimore, Md., 1974; F. Gürsey in Proc. Kyoto Conference on Mathematical Problems in Theoretical Physics, edited by H. Araki (Springer, New York, 1975). See also: R. Casalbuoni, G. Domokos and S. Kövesi-Domokos, Nuovo Cimento A **31**, 423 (1976); **33**, 432 (1976).
[3]M. Günaydin and F. Gürsey, Phys. Rev. D **9**, 3387 (1974) and references quoted there.
[4]In particular, the Günaydin—Gürsey realization of the triality rule violates the superselection rule between hadrons and leptons, see, e.g., P. Sikivie, and F. Gürsey, Phys. Rev. D **16**, 816 (1977). While it has been argued repeatedly that a hadron—lepton superselection rule has no place in a fully unified theory, one may object that this superselection rule is so well satisfied that it may be considered a fundamental law of Nature. Needless to say, it behooves anyone taking this "conservative" point of view to explain the ratios of electric charges of leptons and quarks. We cannot offer a real explanation of charge ratios in this paper; hopefully, that is only due to the fact that flavor groups are not dealt with explicitly here.
[5]A preliminary account of the basic facts has been given in G. Domokos, Johns Hopkins University Report JHU-HET 778 (1977), unpublished. [Invited talk given at the meeting of the AMS, Purdue University, Lafayette (1977).]
[6]H.S. Green, Phys. Rev. **90**, 270 (1953).
[7]O. Klein, J. Phys. Rad. **9**, 1 (1938).
[8]K. Drühl, R. Haag, and J.E. Roberts, Comm. Math. Phys. **18**, 204 (1970).
[9]Y. Ohnuki and S. Kamefuchi, Ann. Phys. (N.Y.) **78**, 64 (1973), and references quoted there.
[10]R.D. Schafer, An Introduction to Nonassociative Algebras (Academic, New York, 1966), p. 141.
[11]R.D. Schafer, Ref. 10, Chap. IV.
[12]R.D. Schafer, Ref. 10, Chap. V.
[13]G. Benkart and J.M. Osborn, private communication (unpublished).
[14]J.R. Faulkner, private communication (unpublished).
[15]G. Domokos and S. Kövesi-Domokos, Johns Hopkins University preprint, JHU-HET 779 (1977), unpublished.
[16]This fact has been independently realized by Rittenberg. (V. Rittenberg, private communication, unpublished).

On the coupling of logics

Antonio Zecca

Istituto di Scienze Fisiche dell'Università, Via Celoria, 16—20133 Milano, Italy
and INFN Sezione di Milano, Italy
(Received 7 June 1977)

A scheme for the coupling of physical systems is proposed within an axiomatic approach to quantum mechanics.

INTRODUCTION

In the logic approach to classical and quantum mechanics, with every physical system there is associated an orthomodular lattice (logic[1,2]). The distributivity or the nondistributivity of the logic depends on the classical or quantum character of the physical system. With the special assumptions leading to Piron's theorem, a logic for a compound quantum system is provided by taking the tensor product logic, namely the standard logic of the tensor product of the Hilbert spaces of the interacting systems. Dealing with classical systems, the logic of a compound classical system can be taken to be the Cartesian product logic, namely, the power set of the Cartesian product of the phase spaces of the interacting systems.

Then the question whether it is possible to give a unified definition of product of logics to cover both cases is of some interest. Since such a definition should be given at the level of orthomodular lattices, it should contain, in principle, also the case of the interaction of classical and quantum systems. This physical situation is relevant for its connection with the problem of quantum measurement in conventional quantum mechanics.[3]

In this note the definitions of *pseudoproduct* and of *product* of logics are proposed. We give necessary and sufficient conditions in order to obtain distributive pseudoproduct logics by coupling distributive logics. For distributive pseudoproduct logics the pseudoproduct is represented by the Cartesian product. We find that the product of irreducible logics is irreducible (and conversely) and that the product logic of standard logics is represented by the tensor product logic. Finally we study an alternative condition for the pseudoproduct of classical and purely quantum logics. This condition, which has a direct physical interpretation makes the pseudoproduct a product and is compatible with the interpretation of the coupling of classical and purely quantum systems as quantum systems endowed with continuous superselection rules. In the three mentioned cases the product satisfies the appropriate requests of existence and uniqueness.

DEFINITIONS AND PRELIMINARY RESULTS

In what follows, by a logic we mean a complete, orthomodular atomistic lattice with the covering property and of length $\geqslant 4$.[4] The least and greatest elements of a logic L will be denoted with $\emptyset$ and $\mathbb{1}$; its atoms with $A(L)$. When x commutes[4] with y, we write xCy. If A is a nonempty subset of L, the complete orthomodular sublattice[4,5] $A'=\{x\in L: xCy\ \forall\ y\in A\}$ of L will be called the commutant of A. The center $C(L)$ of a logic L is such that $C(L)=L'$.[4]

We assume that a logic L corresponds to a physical system Σ. It represents the set of classes of equivalent yes—no experiments on Σ ordered *à la* Piron.[1] If L is distributive $[C(L)=L]$, then L is a classical logic. If L is irreducible $[C(L)=\{\emptyset,\mathbb{1}\}]$, L is a purely quantum logic. If $C(L)$ is nontrivial, we say that L admits superselection rules.[1,2] The three mentioned cases correspond, respectively, to the physical situation in which Σ is a classical or a purely quantum or a quantum physical system.

Definition: A *pseudoproduct* of the logics L, $\tilde{L}$ is a pair $(\bigcirc,\mathcal{L})$, where $\mathcal{L}$ is a logic and $\bigcirc$ is a map $\bigcirc: L\times\tilde{L}\to\mathcal{L}$ such that [we often write $\bigcirc(x,\tilde{x})=x\circ\tilde{x}$]:

P1. $\emptyset\neq x\in L\Rightarrow(x\circ\tilde{x}\leqslant x\circ\tilde{y}\Leftrightarrow\tilde{x}\leqslant\tilde{y})\ \forall\ \tilde{x},\tilde{y}\in\tilde{L}$,

$\tilde{\emptyset}\neq\tilde{x}\in\tilde{L}\Rightarrow(x\circ\tilde{x}\leqslant y\circ\tilde{x}\Leftrightarrow x\leqslant y)\ \forall\ x,y\in L$;

P2. $\mathbb{1}\circ\tilde{\mathbb{1}}=\mathbb{1}_{\mathcal{L}}$; $\emptyset\circ\tilde{x}=x\circ\tilde{\emptyset}=\emptyset_{\mathcal{L}}\ \forall x\in L,\ \forall\ \tilde{x}\in\tilde{L}$;

P3. $\bigcirc(L\times\tilde{L})\equiv L\bigcirc\tilde{L}$ generates $\mathcal{L}$ as a complete orthomodular lattice

P4. $x\circ\tilde{\mathbb{1}}\wedge\mathbb{1}\circ\tilde{x}=x\circ\tilde{x}\ \forall x\in L,\ \forall\ \tilde{x}\in\tilde{L}$;

P5. $(x\circ\mathbb{1})^{\perp}=x^{\perp}\circ\tilde{\mathbb{1}}$; $(\mathbb{1}\circ\tilde{x})^{\perp}=\mathbb{1}\circ\tilde{x}^{\perp}\ \forall x\in L,\ \forall\ \tilde{x}\in\tilde{L}$;

P6. $A(L)\bigcirc A(\tilde{L})\subset A(\mathcal{L})$.

A *product* of L and $\tilde{L}$ is a *pseudoproduct* $(\bigcirc,\mathcal{L})$ such that:

P7. $e(x)\circ e(\tilde{x})=e(x\circ\tilde{x})\ \forall x\in A(L),\ \forall\tilde{x}\in A(\tilde{L})$, $e(x)$ being the central cover[4] of x. The elements of $L\bigcirc\tilde{L}$ will be called *product elements*. (To avoid heavy notations, we do not label, as it has just been done, the lattice operations with reference to the logics in which they work.)

Remark 1: According to Definition 1, the map $x\to x\circ\tilde{x}$ $(\tilde{x}\to x\circ\tilde{x})$ is an isomorphism from L $(\tilde{L})$ onto the sublattice $L\bigcirc\tilde{x}$ $(x\bigcirc\tilde{L})$ of $\mathcal{L}$ when $\tilde{x}\neq\tilde{\emptyset}$ $(x\neq\emptyset)$. The map $\mu(x)=x\circ\tilde{\mathbb{1}}$ $(\tilde{\mu}(\tilde{x})=\mathbb{1}\circ\tilde{x})$ is an ortho-isomorphism from L $(\tilde{L})$ onto the orthocomplemented sublattice $L\bigcirc\tilde{\mathbb{1}}$ $(\mathbb{1}\bigcirc\tilde{L})$ of $\mathcal{L}$.

If $\tilde{L}$ is the logic corresponding to the physical system $\tilde{\Sigma}$, we associate with the compound physical system $\Sigma+\tilde{\Sigma}$ a product $(\bigcirc,\mathcal{L})$ of L and $\tilde{L}$. In the next section it will be seen that this is consistent when any one of Σ, $\tilde{\Sigma}$ is chosen to be a classical or a purely quantum system.

The physical interpretation of the elements of $\mathcal{L}$ is the following. A product element $x\circ\tilde{x}$ corresponds to the yes—no experiments (product tests) on $\Sigma+\tilde{\Sigma}$ ob-

 0022-2488/78/1906-1482$1.00

tained by taking "yes" when both an experiment of x performed on Σ and an experiment of $\tilde{x}$ performed on $\tilde{\Sigma}$ give the answer "yes"; "no" otherwise. The remaining elements of $\mathcal{L}$ are such that none of them contains product tests.

The axioms P1, P2, and P4, P5 express the compatibility between the given interpretation of the product elements and the physical interpretation of the order relations and of the operations existing in each logic. The condition P3 is an irreducibility condition. The request P6 has been done on the analogy of the tensor product of Hilbert spaces. It does not necessary follow from the proposed interpretation of the elements of $\mathcal{L}$. The axiom P7 is a technical condition.

Lemma 1: Let $(\bigcirc, \mathcal{L})$ be a *pseudoproduct* of the logics L, $\tilde{L}$. Then:

(i) $\wedge_{\alpha,\beta}(x_\alpha \circ \tilde{x}_\beta) = \wedge_\alpha x_\alpha \circ \wedge_\beta \tilde{x}_\beta \quad \forall \{x_\alpha\} \subset L, \quad \{\tilde{x}_\beta\} \subset \tilde{L}$;

(ii) $x \circ \tilde{x} \neq \emptyset_{\mathcal{L}} \Rightarrow (x \circ \tilde{x} \leq y \circ \tilde{y} \Longleftrightarrow x \leq y \text{ and } \tilde{x} \leq \tilde{y})$;

(iii) $(x \circ \tilde{x})^\perp = (x^\perp \circ \tilde{x}^\perp) \vee (x^\perp \circ \tilde{x}) \vee (x \circ \tilde{x}^\perp), \quad x \in L, \quad \tilde{x} \in \tilde{L}$.

Proof: (i) From P1 we have $\wedge_\alpha x_\alpha \circ \wedge_\beta \tilde{x}_\beta \leq \wedge_{\alpha,\beta}(x_\alpha \circ \tilde{x}_\beta)$. On the other hand, $\wedge_{\alpha,\beta}(x_\alpha \circ \tilde{x}_\beta) \leq (\wedge_\alpha x_\alpha \circ \tilde{1}) \wedge (1 \circ \wedge_\beta \tilde{x}_\beta) = \wedge_\alpha x_\alpha \circ \wedge_\beta \tilde{x}_\beta$ by using P1, Remark 1, P4. (ii) We show $\emptyset_{\mathcal{L}} \neq x \circ \tilde{x} \leq y \circ \tilde{y} \Rightarrow x \leq y$ and $\tilde{x} \leq \tilde{y}$ since the converse follows easily from P1. By the assumption, (i), and P1 we have $x \circ \tilde{x} = (x \wedge y) \circ (\tilde{x} \wedge \tilde{y}) \leq x \circ (\tilde{x} \wedge \tilde{y})$. Hence $\tilde{x} \leq \tilde{x} \wedge \tilde{y}$ and then $\tilde{x} \leq \tilde{y}$. Analogously, $x \leq y$. (iii) We have $(x \circ \tilde{x})^\perp = (x \circ \tilde{1})^\perp \vee (1 \circ \tilde{x})^\perp = (x^\perp \circ (\tilde{x} \vee \tilde{x}^\perp)) \vee ((x \vee x^\perp) \circ \tilde{x}^\perp) = (x^\perp \circ \tilde{x}) \vee (x^\perp \circ \tilde{x}^\perp) \vee (x \circ \tilde{x}^\perp) \vee (x^\perp \circ \tilde{x}^\perp)$ by using P4, P5, and Remark 1.

The following lemma establishes some immediate properties of the commutant that will be useful in the next section.

Lemma 2: If A, B are subsets of a logic L with $A \cap B$ a nonempty set, then we have:

(i) $A \subset B \Rightarrow B' \subset A'$;

(ii) $A'' \supset A; \quad A''' = A'$;

(iii) $(A \cap B)' \supset (A' \cup B')''$.

THREE APPLICATIONS

We first study the interaction of classical systems.

Proposition 1: Let $(\cap, \mathcal{L})$ be a *pseudoproduct* of the logics L and $\tilde{L}$. Then the following conditions are equivalent:

(i) $\mathcal{L}$ is a Boolean algebra

(ii) $L, \tilde{L}$ are Boolean algebras and $A(L) \bigcirc A(\tilde{L}) \equiv A(\mathcal{L})$.

Proof: (i) $\Rightarrow$ (ii). By the assumptions and Remark 1, $L, \tilde{L}$ are Boolean algebras. If now $x \in A(\mathcal{L})$, $x \notin A(L) \bigcirc A(\tilde{L})$, $\{x_\alpha\} \subset A(L)$, $\{\tilde{x}_\beta\} \subset A(\tilde{L})$, $\vee_\alpha x_\alpha = \mathbb{1}$, $\vee_\beta \tilde{x}_\beta = \tilde{\mathbb{1}}$, then $x = x \wedge \mathbb{1} \circ \tilde{\mathbb{1}} = x \wedge (\vee_\alpha \vee_\beta (x_\alpha \circ \tilde{x}_\beta)) = \vee_\alpha \vee_\beta (x \wedge (x_\alpha \circ \tilde{x}_\beta)) = \emptyset_{\mathcal{L}}$ by developing according to the isomorphism law of Remark 1 and by distributivity.

(ii) $\Rightarrow$ (i). By Huntingtone's theorem[6] it is enough to prove that $A = (A \wedge B) \vee (A \wedge B^\perp) \ \forall A, B \in \mathcal{L}$. Let $A = \vee_\alpha x_\alpha \circ \tilde{x}_\alpha$, $B = \vee_\beta y_\beta \circ \tilde{y}_\beta$ with the $x_\alpha \circ \tilde{x}_\alpha$'s and the $y_\beta \circ \tilde{y}_\beta$'s product atoms. It is not difficult to show that the distributivity of L and $\tilde{L}$ implies that either $x_\alpha \circ \tilde{x}_\alpha \leq B$ or $x_\alpha \circ \tilde{x}_\alpha \leq B^\perp$. Hence $(A \wedge B) \vee (A \wedge B^\perp) \geq (x_\alpha \circ \tilde{x}_\alpha \wedge B) \vee (x_\alpha \circ \tilde{x}_\alpha \wedge B^\perp) = x_\alpha \circ \tilde{x}_\alpha \ \forall \alpha$, and then $(A \wedge B) \vee (A \wedge B^\perp) \geq A$. The converse inequality holds in general.

Remark 2: Let $(\bigcirc, \mathcal{L})$ be *pseudoproduct* of L and $\tilde{L}$. By Proposition 1, if $\mathcal{L}$ is distributive so are L and $\tilde{L}$. The pair $(\bigcirc, \mathcal{L})$ is then a *product* since P7 holds trivially. Moreover, the pair $(\{\mu, \tilde{\mu}\}, \mathcal{L})$ ($\mu, \tilde{\mu}$ being the maps of Remark 1) is a Boolean product of L and $\tilde{L}$ in the sense of Ref. 7. Indeed, by P4, $L \bigcirc \tilde{\mathbb{1}}$ and $\mathbb{1} \bigcirc \tilde{L}$ are independent subalgebras of $\mathcal{L}$. With our assumptions on the lattices, $\mathcal{L}$ is ortho-isomorphic ($\approx$) to $\mathcal{P}(A(\mathcal{L}))$, the power set of its atoms. Analogously $L \approx \mathcal{P}(A(L))$, $\tilde{L} \approx \mathcal{P}(A(\tilde{L}))$. By Proposition 1, Lemma 1 (ii): $\mathcal{L} \approx \mathcal{P}(A(L) \bigcirc A(\tilde{L})) \approx \mathcal{P}(A(L) \times A(\tilde{L}))$. So, in the distributive case, the product function is represented by the Cartesian product function.

Before approaching the purely quantum case and the mixed case we state some results that hold in general. They are obtained at the level of the *pseudoproducts* without making any use of P6.

Proposition 2: Let $(\bigcirc, \mathcal{L})$ be a *pseudoproduct* of the logics L and $\tilde{L}$. Then the following hold:

(i) $(L \bigcirc \tilde{L})'' = \mathcal{L}$;

(ii) $L \bigcirc \tilde{L} \cap (L \bigcirc \tilde{L})' = C(L) \bigcirc C(\tilde{L})$;

(iii) $(C(L) \bigcirc C(\tilde{L}))'' = C(\mathcal{L})$.

Proof: (i) It follows from P3, Lemma 2(ii), and some properties of the commutant.

(ii) Let $\emptyset_{\mathcal{L}} \neq x \circ \tilde{x} \in L \bigcirc \tilde{L} \cap (L \bigcirc \tilde{L})'$. Then, in particular, $x \circ \tilde{x} C y \circ \tilde{\mathbb{1}} \ \forall y \in L$, that is, $x \circ \tilde{x} = ((x \circ \tilde{x}) \wedge (y \circ \tilde{\mathbb{1}})) \vee ((x \circ \tilde{x}) \wedge (y \circ \tilde{\mathbb{1}})^\perp) = ((x \wedge y) \vee (x \wedge y^\perp)) \circ \tilde{x}$, by taking into account P5, Lemma 1(i), and Remark 1. Hence we have $x = (x \wedge y) \vee (x \wedge y^\perp) \ \forall y \in L$. Analogously $\tilde{x} \in C(\tilde{L})$. On the other hand, if $\emptyset_{\mathcal{L}} \neq x \circ \tilde{x} = ((x \wedge y) \vee (x \wedge y^\perp)) \circ ((\tilde{x} \wedge \tilde{y}) \vee (\tilde{x} \wedge \tilde{y}^\perp))$ by expanding according to the isomorphism law of Remark 1, and by Lemma 1

$$\begin{aligned} x \circ \tilde{x} &= ((x \wedge y) \circ (\tilde{x} \wedge \tilde{y})) \vee ((x \wedge y^\perp) \circ (\tilde{x} \wedge \tilde{y})) \\ &\quad \vee ((x \wedge y) \circ (\tilde{x} \wedge \tilde{y}^\perp)) \vee ((x \wedge y^\perp) \circ (\tilde{x} \wedge \tilde{y}^\perp)) \\ &\leq ((x \circ \tilde{x}) \wedge (y \circ \tilde{y})) \vee (x \circ \tilde{x} \wedge (y \circ \tilde{y})^\perp) \ \forall y \circ \tilde{y} \in L \bigcirc \tilde{L}. \end{aligned}$$

Since the converse inequality holds in general, we obtain $x \circ \tilde{x} \in L \bigcirc \tilde{L} \cap (L \bigcirc \tilde{L})'$.

(iii) By taking the commutant in (ii) and by using (i) and Lemma 2, we have $(C(L) \bigcirc C(\tilde{L}))' = \mathcal{L}$. The proof is completed by taking the commutant once again.

Proposition 3: Let $(\bigcirc, \mathcal{L})$ be a *product* of L and $\tilde{L}$. Then $\mathcal{L}$ is irreducible if and only if both L and $\tilde{L}$ are irreducible.

Proof: By Proposition 2(iii), Lemma 2(ii), P2, and Lemma 1(ii), if $C(\mathcal{L}) = \{\emptyset_{\mathcal{L}}, \mathbb{1}_{\mathcal{L}}\}$, then $C(L) = \{\emptyset, \mathbb{1}\}$, $C(\tilde{L}) = \{\tilde{\emptyset}, \tilde{\mathbb{1}}\}$. We show the converse by showing that we have $A(C(\mathcal{L})) = A(C(L)) \bigcirc A(C(\tilde{L}))$ for the atoms of the centers. Indeed $A(C(L)) = \{e(x): x \in A(L)\}$ holds (see Ref. 4, Lemma 10.11), $e(x)$ being the central cover of x. Analogous relations hold for $C(\tilde{L})$ and $C(\mathcal{L})$. By P6, P7, $A(C(\mathcal{L})) \supset A(C(L)) \bigcirc A(C(\tilde{L}))$. If $A \in A(C(\mathcal{L}))$, $A \notin A(C(L)) \bigcirc A(C(\tilde{L}))$ we have the absurd $A = A \wedge \mathbb{1} \circ \tilde{\mathbb{1}} = A \wedge e(\vee_\alpha x_\alpha \circ \vee_\beta \tilde{x}_\beta) = A \wedge (\vee_{\alpha,\beta} e(x_\alpha) \circ e(\tilde{x}_\beta)) = \emptyset_{\mathcal{L}}$, where $\{x_\alpha\} \subset A(L)$,

$\{\tilde{x}_\beta\} \subset A(\tilde{L})$, $\mathbb{1} = \vee_\alpha x_\alpha$, $\tilde{\mathbb{1}} = \vee_\beta \tilde{x}_\beta$. This can be easily seen by the fact that the central cover commutes with the lattice join (Ref. 4, Lemma 5.11) by P7 and by exploiting the distributivity of $C(\mathcal{L})$.

Remark 3: (a) Let M, $\tilde{M}$, $\mathcal{J}$ be logics and $h: M \to L$, $\tilde{h}: \tilde{M} \to \tilde{L}$, $g: \mathcal{L} \to \mathcal{J}$ be ortho-isomorphisms. Then it is easy to check that the map $\bigcirc': M \times \tilde{M} \to \mathcal{J}$ defined by $\bigcirc'(x, \tilde{x}) = g(h(x) \circ \tilde{h}(\tilde{x}))$ is such that $(\bigcirc', \mathcal{J})$ is a *product* of M and $\tilde{M}$.

(b) Let $(\bigcirc, \mathcal{L}(H))$ be a *product* of $\mathcal{L}(\mathfrak{H})$ and $\mathcal{L}(\tilde{\mathfrak{H}})$ [$\mathfrak{H}$, $\tilde{\mathfrak{H}}$, H are Hilbert spaces over the same field of numbers and $\mathcal{L}(\mathfrak{H})$ is the standard logic associated with $\mathfrak{H}$]. Then $H \approx \mathfrak{H} \otimes \tilde{\mathfrak{H}}$.

The need for the study of the coupling of classical and quantum systems and its connection with the problem of measurement interaction has been recalled and stressed in Ref. 3. There the problem has been posed and solved by reinterpreting a classical system as a quantum system with superselection rules. The counterpart of that problem in the context of the logic approach to classical and quantum mechanics can now be addressed by specializing the scheme of the Definition 1.

Proposition 4: Let $(\bigcirc, \mathcal{L})$ be a *pseudoproduct* of the logics L and $\tilde{L}$ with $C(L) = L$, $C(\tilde{L}) = \{\tilde{\emptyset}, \tilde{\mathbb{1}}\}$. Suppose that the following condition holds:

P8. $A \in A(\mathcal{L})$, $x \in A(L)$, and $A \leq x \circ \tilde{\mathbb{1}} \Rightarrow A = x \circ \tilde{x}$ for some $\tilde{x} \in \tilde{L}$.

Then we have:

(i) $A(L) \bigcirc \tilde{\mathbb{1}} = A(C(\mathcal{L}))$;

(ii) $L[\emptyset_{\mathcal{L}}, x \circ \tilde{\mathbb{1}}] = x \bigcirc \tilde{L}$, $(x \in A(L))$, where $L[\emptyset_{\mathcal{L}}, x \circ \tilde{\mathbb{1}}]$ is the segment[4] from $\emptyset_{\mathcal{L}}$ to $x \circ \tilde{\mathbb{1}}$.

Proof: (i) Let $x \in A(L)$ and suppose $A \leq x \circ \tilde{\mathbb{1}}$, $A \in A(C(\mathcal{L}))$. Since $\mathcal{L}$ is atomistic, $A = \vee_\alpha A_\alpha$ with $\{A_\alpha\} \subset A(\mathcal{L})$. There follows $A_\alpha \leq x \circ \tilde{\mathbb{1}}$. By P8, $A_\alpha = x \circ \tilde{x}_\alpha$ and hence $A = x \circ (\vee_\alpha \tilde{x}_\alpha)$. By Proposition 2(ii), $A = x \circ \tilde{\mathbb{1}}$. Hence $A(L) \bigcirc \tilde{\mathbb{1}} \subset A(C(\mathcal{L}))$. The equality can be shown to hold by mimicking the first part of the proof of Proposition 1 with regard to the atoms $A(C(\mathcal{L}))$.

(ii) We have $L[\emptyset_{\mathcal{L}}, x \circ \tilde{\mathbb{1}}] \supset x \bigcirc \tilde{L}$ $(x \in A(L))$. With our assumptions, the sublattice $x \bigcirc \tilde{L}$ of $\mathcal{L}$ is a sublogic of the irreducible logic $L[\emptyset_{\mathcal{L}}, x \circ \tilde{\mathbb{1}}]$. To see this, we have only to check that all works well for the relative orthocomplementation. We have indeed $(x \circ \tilde{x})^\perp \wedge (x \circ \tilde{\mathbb{1}}) = x \circ \tilde{x}^\perp$ by Lemma 1(iii) and by the distributivity that holds for $x \circ \tilde{\mathbb{1}}$ being in $C(\mathcal{L})$. To complete the proof, we show that every atom A of $L[\emptyset_{\mathcal{L}}, x \circ \tilde{\mathbb{1}}]$ is such that $A \in A(L) \bigcirc \tilde{\mathbb{1}}$. Indeed, for such an atom, we have also $A \in A(\mathcal{L})$ (Ref. 4, Lemma 8.18). Since $A \leq x \circ \tilde{\mathbb{1}}$, by P8, P1, $A = x \circ \tilde{x}$ for some $\tilde{x} \in A(\tilde{L})$.

Remark 4: A first consequence of Proposition 4 is that the *pseudoproduct* there studied is also a *product*: From $A(L) \bigcirc \tilde{\mathbb{1}} = A(C(\mathcal{L}))$ we have indeed $e(x \circ \tilde{x}) = x \circ \tilde{\mathbb{1}} = e(x) \bigcirc e(\tilde{x})$, $x \in A(L)$, $x \in A(\tilde{L})$. Another consequence is

$$\mathcal{L} = \oplus (x \bigcirc \tilde{L} : x \in A(L)),$$

that is, $\mathcal{L}$ can be decomposed into the (possibly continuous) direct sum of mutually orthogonal irreducible logics (sectors) each of which is relatively ortho-isomorphic to $\tilde{L}$. This is a consequence of standard results on lattice decomposition theory (Ref. 4, Lemma 10.12) and the fact that x, $y \in A(L)$, $x \neq y$ (and hence $x \leq y^\perp$, L being distributive) imply $x \circ \tilde{x} \leq y^\perp \circ \tilde{x} \leq y^\perp \circ \tilde{\mathbb{1}} \leq (y \bigcirc \tilde{z})^\perp$ $\forall\, \tilde{x}, \tilde{z} \in \tilde{L}$ by P1 and Lemma 1(iii). Accordingly, each $a \in \mathcal{L}$ can be uniquely decomposed into the orthogonal sum $a = \oplus_\alpha a_\alpha$ where the α component is determined by $a_\alpha = a \wedge x_\alpha \bigcirc \tilde{\mathbb{1}}$ $(\{x_\alpha\} \equiv A(L))$. Proceeding as in the first part of the proof of Proposition 4, we have $a_\alpha = x_\alpha \bigcirc \tilde{x}_\alpha(a)$ with $\tilde{x}_\alpha(a) \in \tilde{L}$. With some calculations, the β component of $a^\perp$ is found to be $(a^\perp)_\beta = (\vee_\alpha (x_\alpha \circ \tilde{x}_\alpha(a)))^\perp \wedge x_\beta \circ \tilde{\mathbb{1}} = x_\beta \circ \tilde{x}_\beta^\perp(a)$. Analogously, if $\{a_k\}$ is any family of elements of $\mathcal{L}$ we have $(\vee_k a_k)_\alpha = x_\alpha \circ \tilde{x}_\alpha(\vee_k a_k) = \vee_k (x_\alpha \circ \tilde{x}_\alpha(a_k))$ and $(\wedge_k a_k)_\alpha = \wedge_k (x_\alpha \circ \tilde{x}_\alpha(a_k))$. If now $(\bigcirc, \mathcal{L})$ and $(\bar{\bigcirc}, \bar{\mathcal{L}})$ are *pseudoproducts* of L and $\tilde{L}$ and if the assumptions of Proposition 4 hold, then $\mathcal{L}$ and $\bar{\mathcal{L}}$ are ortho-isomorphic by the map $a = \vee_\alpha (x_\alpha \circ \tilde{x}_\alpha(a)) \to \bar{a} = \bar{\vee}_\alpha (x_\alpha \bar{\circ} x_\alpha(a))$. This is a consequence of the uniqueness of the orthogonal decomposition and of the previous considerations.

The decomposition of $\mathcal{L}$ reduces also the problem of representing the *product* $(\bigcirc, \mathcal{L})$ of Proposition 4 to that of representing irreducible logics. This too is a known fact, at least when the $x \bigcirc \tilde{L}$'s are associated with inner product spaces over the complex numbers (Ref. 4, Sec. 34).

The physical interpretation of P8 is the following. Let us perform a minimal test on the classical system Σ and the test of the existence of the purely quantum system $\tilde{\Sigma}$. Then a minimal test on the compound system $\Sigma + \tilde{\Sigma}$ majorized by the product of the two is itself a product test. This assumption has the consequence that the interaction of a classical system with a purely quantum system can be described in terms of a quantum system endowed with continuous superselection rules. The corresponding superselection sectors are labelled by the atoms of the classical logic, each sector providing an ortho-isomorphic image of the purely quantum logic.

CONCLUDING REMARKS

In this paper we have tried to give an intrinsic characterization of the interaction of physical systems within the logic approach to quantum mechanics. The characterization is based on Definition 1 which requires the atomicity condition for the logics. This condition plays a fundamental role in the deduction of the results of the previous section, but it restricts the class of the physical systems to which the definition applies. So one might ask whether, by relaxing the atomicity condition, one is still able to show the existence and the uniqueness of the product. Another open problem is the study of the Definition 1 in the cases that do not reduce to those treated in the previous section. Finally we observe that Definition 1 can be generalized to describe the coupling of more than two physical systems.

ACKNOWLEDGMENTS

The author thanks A. Frigerio and G. M. Prosperi for discussions on the problem of quantum measure-

ment, V. Gorini for constant interest and for bringing Ref. 3 to his attention, and the referee for useful suggestions.

[1]C. Piron, *Foundations of Quantum Physics* (Benjamin, London, 1976).
[2]V.S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, N.J., 1968), Vol. I.
[3]E.C.G. Sudarshan, Pramana 6, 117 (1976).
[4]F. Maeda and S. Maeda, *Theory of Symmetric Lattices* (Springer-Verlag, Berlin, 1970).
[5]S.S. Holland, Trans. Am. Math.Soc., 108, 66 (1963).
[6]G. Birkhoff, *Lattice Theory* (AMS, Providence, R.I., 1973), Chap. 3, Sec. 10.
[7]R. Sikorski, *Boolean Algebra* (Springer-Verlag, Berlin, 1964), 2nd ed., Chap. 1, Sec. 13.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*