JOURNAL OF MATHEMATICAL PHYSICS

VOLUME 13, NUMBER 7 JULY 1972

Generalized Lienard-Wiechert Fields*

W. E. Couch and E. T. Newman

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 17 January 1972)

We show that a moving point multipole source for a spin-S field gives rise to radiation having (in the rest frame of the source) multipole structure of order S through $2L + S$, where L is the highest order moment of the source.

1. INTRODUCTION

It is well known that in the rest frame of a moving point charge the Lienard–Wiechert radiation fields have a dipole structure. This is one instance of a general property of rest-mass-zero spin-S fields: namely, that a moving multipole source for such fields gives rise to radiation having (in the rest frame of the source) multipole structure of orders S through $2L + S$, where L is the highest order moment of the source. We show how this property arises.

The multipole analysis of the field equations for spin-S fields reveals a certain set of conditions on the source. In the case of the linearized gravitational field ($S = 2$) these conditions form part of the equations of motion[1] of the source. For a spin-2, pole–dipole–quadrupole source these equations of motion do not involve the quadrupole moment.

2. THE SCALAR FIELD

Supposing that the field is due to the motion of a point multipole source we want to obtain the radiation field expressed in the rest frame of the source. This we accomplish by the use of a special coordinate system[2] based on the family of null cones that emanate from the timelike world line of the particle.

Let the world line be given in Minkowski coordinates y^μ by the parametric equations $y^\mu = \xi^\mu(u)$, where the parameter u labels the null cones emanating from the world line and $u = (1/\sqrt{2})\tau$, where τ is the proper time. The Minkowski coordinates of an arbitrary point can be expressed in terms of null coordinates $x^\mu = (u, r, x^2, x^3)$ by

$$y^\mu = \xi^\mu(u) + (r/v)\, l^\mu(x^2, x^3), \tag{2.1}$$

where r is an affine parameter along each null ray l^μ lying on the cones and a measure of the radius of each sphere of constant u and r; $v(u, x^2, x^3) \equiv \dot{\xi}^\mu l_\mu$, where the dot denotes $\partial/\partial u$, v is the velocity $\dot{\xi}^\mu$ projected along l_μ and serves to fix the normalization of l^μ; x^2 and x^3 are related to complex stereographic coordinates ζ and $\bar{\zeta}$ by $\zeta = x^2 + ix^3$. The null vector l^μ sweeps out the directions in the null cone at each $u =$ const and in Minkowski coordinates is given by

$$l^\mu = (\sqrt{2}/4P_0)(1 + \zeta\bar{\zeta}, \zeta + \bar{\zeta}, (\zeta - \bar{\zeta})/i, \zeta\bar{\zeta} - 1), \tag{2.2}$$

where $P_0 = \frac{1}{2}(1 + \zeta\bar{\zeta})$.

In terms of these coordinates the flat-space metric has the form

$$ds^2 = 2[1 - (\dot{v}/v)r]du^2 + 2du dr - (r^2/2P_0{}^2v^2)\, d\zeta d\bar{\zeta}, \tag{2.3}$$

where $\dot{v}$ is the projected acceleration $\dot{v} = \ddot{\xi}^\mu l_\mu$. We denote 2-surface with metric $d\zeta d\bar{\zeta}/v^2\ P_0{}^2$ by S.

The wave equation for a scalar field ϕ takes the form

$$g^{\mu\nu}\phi_{;\mu\nu} = r^2 \frac{\partial\dot{\phi}}{\partial r} + r\dot{\phi} - 2r\left(1 - \frac{\dot{v}}{v}\, r\right)\frac{\partial\phi}{\partial r} - r^2\left(1 - \frac{\dot{v}}{v}\, r\right)\frac{\partial^2\phi}{\partial r^2} - \eth\bar{\eth}\phi = 0. \tag{2.4}$$

The operators $\eth$ and $\bar{\eth}$ are respectively raising and lowering operators for spin-weighted quantities and in general are defined by[3]

$$\eth\eta = 2P^{1-s}\frac{\partial}{\partial\zeta}(P^s\eta), \qquad \bar{\eth}\eta = 2P^{1+s}\frac{\partial}{\partial\bar{\zeta}}(P^{-s}\eta),$$
$$P(u, \zeta, \bar{\zeta}) = P_0 v,$$

η being a quantity having spin weight[4] s. For the scalar field we have simply $\eth\bar{\eth}\phi = 4P^2(\partial^2/\partial\zeta\partial\bar{\zeta})\,\phi$.

The solution of Eq. (2.4) that allows retarded multipoles through order L is

$$\phi = \sum_{n=0}^{L}\frac{f_n(u, \zeta, \bar{\zeta})}{r^{n+1}}, \tag{2.5}$$

where f_n must satisfy

$$(n+1)\dot{f}_{n+1}-(n+1)(n+2)(\dot{v}/v)f_{n+1}=[\eth\bar{\eth}+n(n+1)]f_n, \quad n=L-1,L-2,\dots,0, \quad (2.6)$$

$$\eth\bar{\eth}f_L=-L(L+1)f_L. \quad (2.7)$$

In order to find the multipole structure of the radiation field f_0, we break Eqs. (2.6) and each f_n into their angular parts. The different angular parts are labeled by l and are eigenfunctions of the Laplacian on the unit sphere $\mathcal{S}$. In our coordinate system the equation for the eigenfunctions of the Laplacian is

$$\eth\bar{\eth}\beta_l=-l(l+1)\beta_l. \quad (2.8)$$

These eigenfunctions (essentially linear combinations of spherical harmonics) compose the complete set of functions with which we make multipole expansions. Note that they are not eigenfunctions of the z component of the angular momentum operator.

The acceleration term $\dot{v}/v$ is an eigenfunction of order $l=1$. This may easily be derived by differentiating the curvature of $\mathcal{S}$, i.e., $K=\eth\bar{\eth}\log P=1$, with respect to u and noting that one obtains the eigenvalue equation (2.8) for $l=1$.

Since $\dot{v}/v$ is an eigenfunction of order $l=1$ the quantity $(\dot{v}/v)\beta_l$ contains angular parts of order $l-1$ and $l+1$ (the angular part of order l is not present because β_l has spin weight zero, the part of order $l-1$ is not present when $l=0$).

The angular parts of $\dot{\beta}_l$ are found by differentiation of Eqs. (2.8) and use of the formula[1]

$$\frac{\partial}{\partial u}(\eth\bar{\eth}\eta)=2s\frac{\dot{v}}{v}\eta+2\frac{\dot{v}}{v}\eth\bar{\eth}\eta+s\left(\eth\frac{\dot{v}}{v}\right)\bar{\eth}\eta-s\left(\bar{\eth}\frac{\dot{v}}{v}\right)\eth\eta+\eth\bar{\eth}\dot{\eta} \quad (2.9)$$

in the case $s=0$. We find

$$[\eth\bar{\eth}+l(l+1)]\dot{\beta}_l=2l(l+1)(\dot{v}/v)\beta_l. \quad (2.10)$$

Since $(\dot{v}/v)\beta_l$ has only the angular parts $l-1$ and $l+1$ Eq. (2.10) implies that $\dot{\beta}_l$ has at most the angular parts $l-1, l$, and $l+1$.

Equation (2.7) states that f_L is an eigenfunction of order L (its u dependence may be specified arbitrarily); and so we can begin with Eq. (2.6), $n=L-1$, and proceed to $n=0$ and thereby find which angular parts are present in f_n.

Since $\dot{\beta}_l$ and $(\dot{v}/v)\beta_l$ have orders one step lower and one step higher than β_l we see from Eq. (2.6) that, except for the *lowest* order one, each angular part of f_{n+1} of order l contributes to f_n parts of order $l-1, l$, and $l+1$. It follws that f_n has angular parts of order n through $2L-n$. The highest order part $(2L-n)$ is necessarily present by virtue of the presence of f_L. It can be shown from Eq. (2.10) that the lowest order part $(l=n+1)$ of f_{n+1} does not contribute to Eq. (2.6) at order n; hence the lowest order part of each f_n is not determined by Eq. (2.6) but is an arbitrary function of u that arises from the integration of Eq. (2.6). These arbitrary functions of integration correspond to the additional multipole moments of the source whose orders are less than L.

The radiation field f_0 contains in general 2^l poles from $l=0$ to $l=2L$. If we consider the case when the source has only the single 2^L-pole moment, by a counting argument similar to that above, we find that the multipole range of the radiation field is from $l=1$ to $l=2L$.

3. THE SPIN-S FIELD

The spin-S field, $S>0$, is represented by a (complex) totally symmetric spinor $\phi_{CD\dots K}$ with $2S$ spinor indices and satisfies the rest-mass-zero field equations[5]

$$\nabla^{CM'}\phi_{CD\dots K}=0,$$

where ∇_{CD} denotes spinor covariant differentiation. Denoting $\phi_{CD\cdots K}$ by ϕ_B, $B=0,1,2,\dots,(2S)$, according to $\phi_0\equiv\phi_{0\cdots0}, \phi_1\equiv\phi_{10\cdots0},\dots,\phi_{2S}\equiv\phi_{11\cdots1}$, the field equations in the coordinates associated with the world line of the source become

$$\dot{\phi}_A-\left(1-\frac{\dot{v}}{v}r\right)\frac{\partial\phi_A}{\partial r}-\frac{A+1}{r}\phi_A+(S-A)\frac{\dot{v}}{v}\phi_A=-\frac{1}{r}\eth\phi_{A+1}+A\phi_{A-1}\bar{\eth}\frac{\dot{v}}{v}, \quad (3.1)$$

$$\frac{\partial\phi_{A+1}}{\partial r}+\frac{2S-A}{r}\phi_{A+1}=-\frac{1}{r}\bar{\eth}\phi_A, \quad (3.2)$$

where $A=0,1,\dots,(2S-1)$. The quantities ϕ_A have spin weight $S-A$.

The spin S field that contains retarded multipoles through order L, for $L\geq S$, is given by

$$\phi_B=\sum_{n=2S-B}^{L+S}\frac{f_{B,n}(u,\zeta,\bar{\zeta})}{r^{n+1}}. \quad (3.3)$$

Substitution of Eq. (3.3) into the field equations yields conditions on $f_{B,n}$ that can be shown to reduce to the following:

$$\eth\bar{\eth}f_{0,L+S}=-(L+S)(L-S+1)f_{0,L+S}, \quad (3.4a)$$

$$(2S-m)\dot{f}_{0,m}+(2S-m)(S-m-1)(\dot{v}/v)f_{0,m}=[\eth\bar{\eth}-(m-1)(2S-m)]f_{0,m-1}, \quad (3.4b)$$

$$m=(2S+1),\dots,(L+S), \quad f_{A,L+S+1}=f_{A,2S-A-1}=0,$$

$$\dot{f}_{A,2S-A}-(S+1)(\dot{v}/v)f_{A,2S-A}=-\eth f_{A+1,2S-A-1}, \quad (3.4c)$$

$$(2S-A-n-1)f_{A+1,n}=-\bar{\eth}f_{A,n}, \quad n=(2S-A),\dots,(L+S). \quad (3.4d)$$

We now find the multipole structure of the field.[6] The eigenfunctions of spin weight s satisfy

$$\eth\bar{\eth}\beta_l=-(l+s)(l-s+1)\beta_l$$

and $(\dot{v}/v)\beta_l$ and $\dot{\beta}_l$ have angular parts $l-1$, l and $l+1$. (The part of order $l-1$ is not present when $l=s$). Equation (3.4a) states that $f_{0,L+S}$ is an eigenfunction of order L; hence we can begin with $f_{0,L+S}$ and proceed through Eqs. (3.4b), (3.4c), and (3.4d) to obtain the multipole structure of $f_{B,n}$. The analysis of Eqs. (3.4) is done in the same manner as it was for Eqs. (2.6) and (2.7). We find $f_{0,n}$ has multipoles of order $n-S$ to order $(2L+S-n)$, and $f_{B,2S-B}$ has multipoles of order $|S-B|$ to order $(2L-S+B)$. In particular the radiation field $f_{2S,0}$ has multipoles of order S to order $2L+S$.

It can be shown that the lowest order multipole of each of the quantities $f_{0,n}$ and $f_{B,2S-B}$ for $B \leq S$ is not determined by Eqs. (3.4). These are the multipole moments of the source, they may be given arbitrarily. The highest order part of each $f_{0,n}$ and $f_{B,2S-B}$, $B \leq S$, is necessarily present due to the presence of the 2^L-pole.

In the case of a source that has only the single 2^L-pole the counting argument applied to Eqs. (3.4) shows that the radiation field contains multipoles of order S to order $2L + S$. The beginning order for the radiation field due to a 2^L-pole source is established only by a counting argument, and so it is conceivable that in special cases cancellations would occur such that the lowest multipole order of the radiation field would be greater than S.

For the case $L < S$ we must have $\phi_B = 0$ for $B \leq S - L - 1$. We find that the radiation field for this case also has multipoles of order S to order $2L + S$.

These results may be more clearly understood by referring to the array $f_{B,n}$ shown in Table I. The term $f_{0,L+S}$ in the upper right corner of the chart has the single multipole L according to Eq. (3.4a). Equation (3.4b) shows that each successive term to the left of $f_{0,L+S}$ has multipoles that range from one order lower to one order higher than its preceeding term. Then Eq. (3.4c) shows that in general the multipole range continues to change in this manner as we proceed down the diagonal from $f_{0,2S}$ to $f_{S,S}$. From $f_{S,S}$ to $f_{2S,0}$ the lowest order and the highest order multipoles increase by one with each successive term. Equation (3.4d) shows that in each column every term below the top is determined by the term immediately above it, as indicated by the vertical arrows.

TABLE I.

	$\frac{1}{r}$	$\cdots$		$\frac{1}{r^{2S+1}}$	$\cdots$	$\frac{1}{r^{L+S+1}}$
ϕ_0	0	$\cdots$	0	$f_{0,2S}$ ←	$\cdots$	← $f_{0,L+S}$
			↙	↓	↓↓↓	↓
ϕ_1	0	$\cdots$	0 $f_{1,2S-1}$			$f_{1,L+S}$
			↙ ↓			↓
$\vdots$	$\vdots$	$\cdot$	$\cdot$			$\vdots$
						↓
ϕ_S	0	0 $f_{S,S}$	$\cdots$			$f_{S,L+S}$
		↙ ↓				↓
$\vdots$	$\vdots$	$\cdot$				$\vdots$
	0 ↙					↓
ϕ_{2S}	$f_{2S,0}$		$\cdots$			$f_{2S,L+S}$

A nonsingular spin weighted quantity can have no angular parts such that $l < |s|$. In particular, when $A \geq S$, $f_{A+1,2S-A-1}$ can have no angular parts for which $l < |S - A - 1| = A - S + 1$; hence (for $A \geq S$) the left-hand side of Eqs. (3.4c) can have no angular parts for which $l = A - S - 1$ or $l = A - S$. (The first condition is not present when $A = S$). These conditions constitute restrictions on the source properties. In electromagnetism ($S = 1$), for example, these conditions are merely charge conservation.

In linearized gravitational theory ($S = 2$) these conditions compose part of the equations of motion of the source. Newman and Young[1] have exhibited these conditions explicitly for the case of a pole-dipole source and have obtained the equations of motion. We have found that when the source is taken to have, in addition, an intrinsic quadrupole moment the conditions that result from Eqs. (4.3c) do not involve the quadrupole moment.

* Research supported in part by the National Science Foundation.

1 E. T. Newman and R. Young, J. Math. Phys. **11**, 3154 (1970).
2 A. Held, E. T. Newman, and R. Posadas, J. Math, Phys. **11**, 3145 (1970).
3 J. N. Goldberg, *et al.*, J. Math. Phys. **8**, 2155 (1967).
4 The spin weight s, of a quantity is defined in terms of the quantity's transformation properties under rotations and is not to be confused with the spin (S) of the field; the spin-S field involves scalars having spin weights from $-S$ to S.
5 See, for example, R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).
6 The electromagnetic field ($S = 1$) of moving multipoles has been found by J. R. Ellis, J. Math. Phys. **7**, 1185 (1966), using methods different from ours.

Closed Rotating Cosmologies Containing Matter Described by the Kinetic Theory. Entropy Production in the Collision Time Approximation*

Richard A. Matzner
Department of Physics, University of Texas, Austin, Texas 78712
(Received 13 October 1971; Revised Manuscript Received 21 February 1972)

A collision time approximation for the production of partly collisional particles ("neutrinos") from a collision dominated fluid ("electrons and photons") within closed anisotropic Bianchi type IX cosmologies is discussed. Two distinct, constant, times are introduced; the neutrino production rate t_e^{-1} is distinct from the neutrino destruction rate t_ν^{-1}. The conservation of energy and momentum is discussed, and it is shown, for a subclass of rotating closed cosmologies and for Maxwell–Boltzmann neutrinos, that the local entropy production rate $s^\mu{}_{;\mu}$ is positive, aside from "rest entropy" terms, i.e., terms representing a fixed amount of entropy permanently attached to each neutrino (which goes out of existence when the neutrino is destroyed and reappears when a neutrino is regenerated). The results here contain as appropriate limits the same conclusions within special relativity and within simpler (Bianchi type I) cosmologies.

1. INTRODUCTION

In previous papers[1,2] (referred to hereafter as IX-A and IX-B), we have given a discussion of kinetic theory in closed rotating cosmologies of Bianchi type IX.[3] These models are rotating, anisotropic generalizations of the closed Robertson–Walker cosmologies.[4,5] In papers IX-A and IX-B, we considered anisotropic world models containing a mixture of fluid constituents (say a thermalized photon–electron–positron gas) and a collection of collisionless particles (say neutrinos) which were presumably produced at some instant in the evolution of the model.

It can be shown that the lowest order multipole of each of the quantities $f_{0,n}$ and $f_{B,2S-B}$ for $B \le S$ is not determined by Eqs. (3.4). These are the multipole moments of the source, they may be given arbitrarily. The highest order part of each $f_{0,n}$ and $f_{B,2S-B}$, $B \le S$, is necessarily present due to the presence of the 2^L-pole.

In the case of a source that has only the single 2^L-pole the counting argument applied to Eqs. (3.4) shows that the radiation field contains multipoles of order S to order $2L + S$. The beginning order for the radiation field due to a 2^L-pole source is established only by a counting argument, and so it is conceivable that in special cases cancellations would occur such that the lowest multipole order of the radiation field would be greater than S.

For the case $L < S$ we must have $\phi_B = 0$ for $B \le S - L - 1$. We find that the radiation field for this case also has multipoles of order S to order $2L + S$.

These results may be more clearly understood by referring to the array $f_{B,n}$ shown in Table I. The term $f_{0,L+S}$ in the upper right corner of the chart has the single multipole L according to Eq. (3.4a). Equation (3.4b) shows that each successive term to the left of $f_{0,L+S}$ has multipoles that range from one order lower to one order higher than its preceeding term. Then Eq. (3.4c) shows that in general the multipole range continues to change in this manner as we proceed down the diagonal from $f_{0,2S}$ to $f_{S,S}$. From $f_{S,S}$ to $f_{2S,0}$ the lowest order and the highest order multipoles increase by one with each successive term. Equation (3.4d) shows that in each column every term below the top is determined by the term immediately above it, as indicated by the vertical arrows.

TABLE I.

$$
\begin{array}{c|ccccccc}
 & \frac{1}{r} & \cdots & & & \frac{1}{r^{2S+1}} & \cdots & \frac{1}{r^{L+S+1}} \\
\hline
\phi_0 & 0 & \cdots & & 0 & f_{0,2S} \ \leftarrow & \cdots & \leftarrow f_{0,L+S} \\
 & & & & \swarrow & \downarrow & \downarrow\downarrow\downarrow & \downarrow \\
\phi_1 & 0 & \cdots & 0 & f_{1,2S-1} & & & f_{1,L+S} \\
 & & & \swarrow & \downarrow & & & \downarrow \\
\vdots & \vdots & & \cdot\,\cdot\,\cdot & & & & \vdots \\
 & & & & & & & \downarrow \\
\phi_S & 0 & 0\ f_{S,S} & & \cdots & & & f_{S,L+S} \\
 & & \swarrow\ \downarrow & & & & & \downarrow \\
\vdots & \vdots & \cdot\,\cdot\,\cdot & & & & & \vdots \\
 & 0\ \swarrow & & & & & & \downarrow \\
\phi_{2S} & f_{2S,0} & & & \cdots & & & f_{2S,L+S}
\end{array}
$$

A nonsingular spin weighted quantity can have no angular parts such that $l < |s|$. In particular, when $A \ge S$, $f_{A+1,2S-A-1}$ can have no angular parts for which $l < |S - A - 1| = A - S + 1$; hence (for $A \ge S$) the left-hand side of Eqs. (3.4c) can have no angular parts for which $l = A - S - 1$ or $l = A - S$. (The first condition is not present when $A = S$). These conditions constitute restrictions on the source properties. In electromagnetism ($S = 1$), for example, these conditions are merely charge conservation.

In linearized gravitational theory ($S = 2$) these conditions compose part of the equations of motion of the source. Newman and Young[1] have exhibited these conditions explicitly for the case of a pole-dipole source and have obtained the equations of motion. We have found that when the source is taken to have, in addition, an intrinsic quadrupole moment the conditions that result from Eqs. (4.3c) do not involve the quadrupole moment.

---

* Research supported in part by the National Science Foundation.

1 E. T. Newman and R. Young, J. Math. Phys. **11**, 3154 (1970).

2 A. Held, E. T. Newman, and R. Posadas, J. Math, Phys. **11**, 3145 (1970).

3 J. N. Goldberg, *et al.*, J. Math. Phys. **8**, 2155 (1967).

4 The spin weight s, of a quantity is defined in terms of the quantity's transformation properties under rotations and is not to be confused with the spin (S) of the field; the spin-S field involves scalars having spin weights from $-S$ to S.

5 See, for example, R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

6 The electromagnetic field ($S = 1$) of moving multipoles has been found by J. R. Ellis, J. Math. Phys. **7**, 1185 (1966), using methods different from ours.

Closed Rotating Cosmologies Containing Matter Described by the Kinetic Theory. Entropy Production in the Collision Time Approximation*

Richard A. Matzner

Department of Physics, University of Texas, Austin, Texas 78712

(Received 13 October 1971; Revised Manuscript Received 21 February 1972)

A collision time approximation for the production of partly collisional particles ("neutrinos") from a collision dominated fluid ("electrons and photons") within closed anisotropic Bianchi type IX cosmologies is discussed. Two distinct, constant, times are introduced; the neutrino production rate t_e^{-1} is distinct from the neutrino destruction rate t_ν^{-1}. The conservation of energy and momentum is discussed, and it is shown, for a subclass of rotating closed cosmologies and for Maxwell-Boltzmann neutrinos, that the local entropy production rate $s^\mu{}_{;\mu}$ is positive, aside from "rest entropy" terms, i.e., terms representing a fixed amount of entropy permanently attached to each neutrino (which goes out of existence when the neutrino is destroyed and reappears when a neutrino is regenerated). The results here contain as appropriate limits the same conclusions within special relativity and within simpler (Bianchi type I) cosmologies.

1. INTRODUCTION

In previous papers[1,2] (referred to hereafter as IX-A and IX-B), we have given a discussion of kinetic theory in closed rotating cosmologies of Bianchi type IX.[3] These models are rotating, anisotropic generalizations of the closed Robertson-Walker cosmologies.[4,5] In papers IX-A and IX-B, we considered anisotropic world models containing a mixture of fluid constituents (say a thermalized photon-electron-positron gas) and a collection of collisionless particles (say neutrinos) which were presumably produced at some instant in the evolution of the model.

At no time did we discuss dissipative mechanisms in such a collection of particles. But a collection of electrons certainly interacts (although weakly) with neutrinos that pass sufficiently close, so that one is forced, on this picture, to consider collisional systems.

We shall take the simplest possible approach to the kinetic theory of occasionally colliding particles; we will introduce a collision time approximation.

The method here will be one suggested by Misner and used by Matzner and Misner[6] and Matzner.[7] These previous papers dealt with a Bianchi type I cosmology and will be hereafter denoted I-I and I-II, respectively. The type I models have infinite spatial sections (as opposed to the compact 3-surfaces in type IX), and furthermore the geometry of the model requires[8] $T_{0i} = 0$. This means that there can be no transport of matter as observed from the frame of an observer to whom the universe looks homogeneous. Furthermore, the net rotation of fluids in such a type I model must vanish.

This type IX model uses the techniques first applied to type I models in I-I and I-II. Because of the additional complexity allowed in type IX models, we shall find that many of the simple-minded techniques applied in type I will not work here. In general, the difference is that more attention to thermodynamics is demanded by type IX models. These calculations are interesting both as they pertain to cosmology and as simple examples of nonequilibrium processes. In fact, important cosmological questions such as the dissipation of anisotropy are not yet resolved in these more difficult type IX models.

The central result of this paper is that, according to the generalized collision-time approximation used here, entropy increases for Maxwell–Boltzmann neutrinos within a particular subset of type IX models. It contains, therefore, the proof of this fact both within the framework of special relativity and within the type I dissipation models previously discussed. We expect on physical ground that the same result must hold for other statistics, but the calculations are much more complicated in that case and we have not carried them out. The calculation for general type IX also involves much additional complexity and that investigation is also left for the future.

We may compare the methods used here to other approaches, such as those of Stewart. Here we approximate the Boltzmann equation (by a collision time approximation) which gives reasonable results over a wide range of behavior—from collisional to essentially collisionless particles. We treat the dynamics of the geometry in these closed universes exactly. Stewart has taken two approaches which complement this one. He and others have given a treatment[9,10] which approximates certain dissipative effects by their upper bounds in the long mean-free time limit. The stress tensor then is not directly computed from the distribution of particles in the model. Hence there is also an approximation in the treatment of the dynamics. On the other hand, he has also (with Anderson) given a discussion of transport theory, calculating viscosity and thermal conductivity for instance. For a summary, see Ref. 11. These results are applicable to cosmologies in the very short mean-free time regime. Our (different) approximation allows us to bridge these two regimes and to obtain a somewhat more accurate treatment of the dynamics in the almost collisionless regime.

2. METRIC AND STATEMENT OF PROBLEM

The metric describing a rotating type IX model has been given many times previously[1,12,13]

$$ds^2 = -dt^2 + R^2(t)e^{2\beta}_{ij}\sigma^i\sigma^j \tag{2.1}$$

Here the σ^i obey the type IX curl relation $d\sigma^i = \epsilon_{ijk}\sigma^j\sigma^k$. In Eq. (2.1), $R(t)$ is a function giving the radius of the universe, and β_{ij} is a traceless 3×3 matrix. We introduce the notation

$$\alpha = \ln R/R_{\max} \quad (= -\Omega + \text{const}), \tag{2.2}$$

where $R_{\max}$ is the maximum radius of the universe. Notice α increases as R increases, in contrast to Misner's[14] "Ω-time" defined as $-\alpha + \text{const}$. The usage α is consistent with the type I discussion given by Misner[8] and also found in I-I and I-II. It was also used for type IX by MSW.[13] We shall use α as a time because we shall always treat the early expanding phase of the universe, and α is then a single-valued function of time and can be used as a perfectly good time variable. The sign of α in Eq. (2.2) above is chosen because we expect the α-time to increase as the entropy increases while the universe is emerging from its initial singularity.

The matrix β_{ij} appearing in (2.1) can be partly parametrized by giving two parameters that depend only on its eigenvalues.[8]

$$\beta_+ = -\tfrac{1}{2}\beta_3, \quad \beta_- = (\beta_1 - \beta_2)/(2\sqrt{3}), \tag{2.3}$$

with $\beta_1, \beta_2, \beta_3$ the eigenvalues of β taken in any order. A complete parametrization of β requires a specification also of the orientation of its principal axes. If the matrix $b_{ij} = \mathrm{diag}(\beta_1, \beta_2, \beta_3)$, we define three Euler angles φ, ψ, θ by

$$\beta = e^{-\psi\kappa_1}e^{-\theta\kappa_2}e^{-\varphi\kappa_1}b\,e^{\varphi\kappa_1}e^{\theta\kappa_2}e^{\psi\kappa_1}, \tag{2.4}$$

where the matrices κ_1 and κ_2 are the generators of rotation in the (1, 2) and (2, 3) plane, respectively.[1,12]

In a general type IX model, all of the set $(\beta_+, \beta_-, \varphi, \psi, \theta)$ are functions of time. The $G_{0i} = T_{0i}$ equations show that the momentum flux T_{0i} vanishes only if $\varphi = \text{const}$, $\psi = \text{const}$, $\theta = \text{const}$, and a transformation can then be performed to set them all to zero.[15,16] If $T_{0i} \neq 0$, there is rotation of the basis tetrad compared to parallel transport along the t-lines, and also the average kinematical rotation[17] is nonzero.[1,13,15]

There is an important subclass of models which have nondiagonal β_{ij} (hence, rotate). This type has been called nontumbling (MSW)[13] because in this case the principal axes of the metric rotate about a fixed direction, which may be taken as the 3 axis. The metric components g_{13} and g_{23} then vanish, and the only nondiagonal component is g_{12}. We shall not write the metric out at this time; see IX-A or Ref. 15.

In nontumbling models, only the momentum flux component T_{03} is nonzero. For simplicity in this paper, we will deal only with the nontumbling rotating models. We shall take models, therefore, in which each constituent species has a momentum flux only along the

3 axis (the rotation axis). We feel that this case will give all the essential physics, and it certainly leads to simpler mathematical expressions.

As we mentioned in Sec. 1, we wish to deal with a situation in which there are at least two types of matter. The first type we take to be a collision dominated fluid; an example of such a fluid is a collection of photons, positrons, and electrons at temperatures of order 10^{10} K. The electromagnetic interactions of such particles assure that the mean free time between collisions in such a system is very short so that it behaves as a fluid. It thus has a rest frame with a well-defined 4-velocity u^α (defined by either the requirement of zero number flux or equivalently by the requirement of zero momentum flux; they are the same in very collisional systems). It also has a well-defined energy density ${}_f\rho$ and a pressure ${}_fp$ which is isotropic in the rest frame and which is taken as ${}_fp = {}_f\rho/3$ since we deal with very hot, very relativistic systems.

The other matter in the models we consider is a collection of occasionally colliding (but otherwise free) particles which must be described by use of kinetic theory. We thus postulate the existence of a homogeneous distribution function[18] F. In terms of this distribution function F, the stress tensor arising from the almost collisionless particles is

$$T_{\mu\nu} = R^{-3}\int F q_\mu q_\nu (q^0)^{-1} d^3q. \tag{2.5}$$

Here the q_α are the components of the momentum of a particular one of the almost collisionless particles expressed in the $\{dt, \sigma^i\}$ basis of Eq. (2.1), and $d^3q = dq_1 dq_2 dq_3$. In the following discussion, it will be useful to have explicitly calculated the neutrino stress tensor. We shall follow IX-A in this analysis. We assume the neutrino distribution function represents particles which were in a thermal distribution at the instant α^d and subsequently evolved freely "decoupled." The 4-velocity defined by this thermal distribution is u^α. In what follows, a superscript zero (^{0}) above a symbol indicates its components in an orthonormal frame $\{dt, \omega^i\}$ with $\omega^i = Re^\beta_{ij}\sigma^j$. Hence, $\overset{0}{u}{}^\alpha$ are the components of u^α in this frame.

Because we take a distribution function which was thermal at the instant α^d, F was a function of the form $f(u_\alpha q^\alpha/T_d)$ at that instant. (T_d is the decoupling temperature.)

Subsequent to the instant α^d, the distribution function is assumed to evolve according to the collisionless Boltzmann equation and hence can be expressed in terms of the constants of geodesic motion. We shall use the quantities $k_i = q_i(\alpha^d)$, the 3-momentum components at the decoupling time α^d. We shall always, in what follows, consider a collisionless distribution function to be a function of these k_i. In that case such a distribution function is *not* a function of time, and all time derivatives of the distribution function vanish.

Let us compute T^{00} for a distribution function which was thermal at some instant:

$$T^{00} = R^{-3}\int (R^{-2}e^{-2\beta}_{ij} q_i q_j)^{1/2} f(R_d^{-1}T_d^{-1}k[\ ])d^3q. \tag{2.6}$$

Here $k = (k_l k_l)^{1/2}$ is a constant of the motion (IX-A).

The factor in the argument of f which is not written out in (2.6) is

$$[\ ] = \overset{0}{u}{}^0 (e^{-2\beta^d}_{ij} m_i m_j)^{1/2} - \overset{0}{u}_l e^{-\beta^d}_{lk} m_k. \tag{2.7}$$

In Eq. (2.7), $m_i = k_i/k$. The argument of f in (2.6) and (2.7) is simply $u_\alpha k^\alpha/T_d$ expressed in the $\{dt, \sigma^i\}$ frame.

In IX-A, it is shown that the Jacobian connecting the variables q_i with their initial values k_i is unity, as is the Jacobian connecting the angular element formed from the angles m_i with that formed from the angles $n_i = q_i/(q_l q_l)^{1/2}$. The integration element in (2.6) can then be written $k^2 dk d\omega$; a factor $k = q$ can be factored from the q^0 term. Then a substitution $z = k[\ ]$ $R_d^{-1}T_d^{-1}$ in the k integral gives

$$T^{00} = R^{-4}R_d^4 T_d^4 \int d\omega \frac{(e^{-2\beta}_{ij} n_i n_j)^{1/2}}{4\pi[\ ]^4} \int 4\pi f(z) z^3 dz \tag{2.8}$$

Since f is a well-defined function, the z integration gives a constant which we call a_ν. It is the "Stefan-Boltzmann" constant appropriate to the type of particle under consideration. The angular integration yields an object which depends on the present anisotropy β, the decoupling anisotropy β^d, and the values $\overset{0}{u}{}^\alpha$ giving the rest frame of the matter when it decoupled. It also depends on the metric behavior in the interval between α^d and α in a highly nonlinear way since the metric affects the evolution of the n_i, which for the purposes of integration are considerd functions of the m_i and of t. We write[1]

$$T^{00} = R^{-4}R_d^4 T_d^4 a_\nu [1 + V_\nu(\beta, t)], \tag{2.9}$$

where $V_\nu \geq 0$ is defined by the angular integral in (2.8).

Similarly, we can proceed to compute the momentum flux T_{0i}. We have

$$\begin{aligned} T_{0i} &= -R^{-3}\int q_i f(R_d^{-1}T_d^{-1}k[\ ])d^3q \\ &= -R^{-3}R_d^4 T_d^4 \int n_i (d\omega/4\pi)[\ ]^{-4}\int 4\pi f(z) z^3 dz \\ &= -R^{-3} a_\nu R_d^4 T_d^4 l_i, \end{aligned} \tag{2.10}$$

where l_i is defined by the angular integral

The components T_{ij} can be obtained as in IX-A by noting that $T^{00} = T_k^k$, while the anisotropic stresses may be obtained from the expression (IX-A)

$$\overset{0}{T}_{ij} - \tfrac{1}{3}\delta_{ij}\overset{0}{T}_{kk} = -\rho_\nu \frac{\partial V_\nu(\beta, t)}{\partial \beta_{ij}}. \tag{2.11}$$

The derivation of Eqs. (2.6)-(2.11) is more thoroughly discussed in IX-A.

3. THE COLLISION TIME APPROXIMATION

The evolution of the distribution function F is given by the Boltzmann equation.[18] However, the Boltzmann equation is a highly nonlinear integro-differential equation and hence is quite difficult to solve. We shall instead use a modified form of the collision time approximation.

Consider the following situation. At some instant α' in the evolution of the universe, the fluid (collision-dominated) matter in the model has 4-velocity $u^\gamma(\alpha')$ and a temperature $T(\alpha')$. Let us suppose that, at the instant α', there is suddenly produced a thermal distribution of particles, in the rest frame of the fluid, with the same temperature as the fluid. That is, at the instant α', the distribution function for these particles is isotropic in the rest frame defined by $u(\alpha')$ and is in all respects a thermal distribution with temperature $T(\alpha')$. Let us further suppose that, subsequently to the instant α', the distribution function evolves according to the collisionless Boltzmann equation, e.g., the particles are thermally produced neutrinos. We label such a function, which was thermally decoupled at α', $f_{\alpha'}$. In a system in which the "neutrinos" were produced over an interval of time, we expect the total distribution function to be a superposition of such $f_{\alpha'}$,

$$F(\alpha) = \int_{-\infty}^{\alpha} d\alpha' f_{\alpha'} W(\alpha, \alpha'). \tag{3.1}$$

Here $W(\alpha, \alpha')$ is a nonnegative weighting function.

The evolution of the distribution of occasionally colliding particles is postulated, as was done by Matzner,[7] by giving the time evolution of the weighting function $W(\alpha, \alpha')$. We postulate

$$\frac{\partial W_\nu}{\partial \alpha} = -\frac{W_\nu}{\dot{\alpha} t_\nu} + \frac{\delta(\alpha - \alpha')}{\dot{\alpha} t_e} \tag{3.2}$$

(the dot means d/dt). Here t_ν is the (average) time required for collisions to remove neutrinos from the system (e.g., the time required for a neutrino–antineutrino pair to collide and produce an $e^\pm$ pair), while t_e is the average time for production of neutrinos from the electron fluid. Both t_ν and t_e realistically depend on parameters entering the system, but here we take them to be constants.

The meaning of Eq. (3.2) becomes clear when we compute the time derivative of the distribution function defined by Eq. (3.1), using Eq. (3.2). Making the assumption that $W_\nu(\alpha, \alpha')$ vanishes for all $\alpha' \geq \alpha$, we find

$$\frac{\partial F}{\partial t} = -\int \frac{W_\nu}{t_\nu} f_{\alpha'} d\alpha' + \frac{F_t}{t_e} \tag{3.3}$$

$$= -\frac{F}{t_\nu} + \frac{F_t}{t_e}. \tag{3.4}$$

Hence, F satisfies a collision time approximation with distinct production and destruction times. In Eqs. (3.3) and (3.4), F_t is a thermal distribution which has the same rest frame and temperature as the electrons at the current instant.

4. THE CONSERVATION LAWS

Equations (3.3) and (3.4) are the only statements we shall have to make about the material interactions, aside from the equation ${}_f p = \frac{1}{3}{}_f\rho$ assumed for the electron fluid. All the other information is contained within the conservation laws. One of them,

$$T^{0\mu}{}_{;\mu} = 0,$$

reads (IX-A)

$$(T^{00}R^4)^{\cdot} R^{-4} = -\overset{0}{T}_{ij}\sigma_{ij} \tag{4.1}$$

Here[8]

$$\sigma_{ij} \equiv \tfrac{1}{2}[(e^\beta)^{\cdot} e^{-\beta} + e^{-\beta}(e^\beta)^{\cdot}]_{ij}. \tag{4.2}$$

Because β is traceless, σ_{ij} is also traceless. Defining ${}_fT^{\mu\nu} = ({}_f\rho + {}_fp)u^\mu u^\nu + {}_fp g^{\mu\nu}$, we have

$$R^{-4}({}_fT^{00}R^4)^{\cdot} + \sigma_{ij}({}_f\rho + {}_fp)v_i v_j + (\rho_\nu[1 + \langle V\rangle]R^4)^{\cdot} R^{-4} = \sigma_{ij}\rho_\nu \left\langle \frac{\partial V}{\partial \beta_{ij}} \right\rangle. \tag{4.3}$$

Here $v^i \equiv v_i = R^{-1} e^{-\beta}_{ij} u_j$ is introduced as a shorthand for $\overset{0}{u}_i$. The angular brackets denote an average; for any quantity $A(\alpha')$ which is a property of a particular distribution which is decoupled at the instant α',

$$\langle A \rangle \equiv \frac{\int_{-\infty}^{\alpha} (R'T'/R)^4 W_\nu(\alpha, \alpha') A(\alpha') d\alpha'}{\int_{-\infty}^{\alpha} (R'T'/R)^4 W_\nu(\alpha, \alpha') d\alpha'}. \tag{4.4}$$

Defining the quantity ρ_ν introduced in Eq. (4.3),

$$\rho_\nu = a_\nu \int_{-\infty}^{\alpha} (R'T'/R)^4 W_\nu(\alpha, \alpha') d\alpha', \tag{4.5}$$

we rewrite Eq. (4.4):

$$\langle A \rangle \equiv \frac{a_\nu \int_{-\infty}^{\alpha} W_E(\alpha, \alpha') A(\alpha') d\alpha'}{\rho_\nu R^4} \tag{4.6}$$

with

$$W_E = (R'T')^4 W_\nu. \tag{4.7}$$

When forming the averages involving $V_\nu(\beta, t)$, we have suppressed the subscript ν. Notice that the right side of Eq. (4.3) uses Eq. (2.11).

Now

$$\begin{aligned} \frac{d}{d\alpha}(\rho_\nu R^4[1 + \langle V\rangle]) &= \frac{d}{d\alpha}\left(a_\nu \int (1 + V_\nu) W_E d\alpha'\right) \\ &= a_\nu \int \left[-\frac{W_E}{\dot{\alpha} t_\nu} + (R'T')^4 \frac{\delta(\alpha - \alpha')}{\dot{\alpha} t_e}\right](1 + V_\nu) d\alpha' \\ &\quad + \sigma_{ij} a_\nu \int \frac{\partial V_\nu}{\partial \beta_{ij}} W_E d\alpha' \end{aligned} \tag{4.8}$$

(Because V_ν depends on time through n_i as functions of m_j and t as well as through β, it might be expected that other terms would appear in $dV_\nu/d\alpha$. However, it is shown in IX-A that such terms in fact vanish.)

The last term in Eq. (4.8) is the same as the right-hand side of Eq. (4.3). Furthermore,

$$a_\nu (RT)^4 [1 + V_\nu] |_{\alpha^d = \alpha},$$

which is a factor in the result of the δ-function integration on the right in Eq. (4.8), is the energy density of a thermal distribution of neutrinos with the same 4-velocity and temperature as the electron fluid at the instant α. Hence, aside from statistics, the energy density must equal that of an electron gas with this temperature and velocity:

$$a_\nu (RT)^4 [1 + V_\nu] |_{\alpha^d = \alpha} = [({}_f\rho + {}_fp)(u^0)^2 - {}_fp] rR^4,$$

where the over-all factor r is inserted to take account of differences in statistics.

Equation (4.8), therefore, reads

$$\begin{aligned} \frac{d}{d\alpha}(\rho_\nu R^4[1 + \langle V\rangle]) &= -\frac{\rho_\nu R^4[1 + \langle V\rangle]}{\dot{\alpha} t_\nu} \\ &\quad + \sigma_{ij} a_\nu \int \frac{\partial V_\nu}{\partial \beta_{ij}} W_E d\alpha' \\ &\quad + \frac{({}_f\rho + {}_fp)(u^0)^2 - {}_fp}{\dot{\alpha} t_e} rR^4, \end{aligned} \tag{4.9}$$

and the conservation law, Eq. (4.1), reads

$$\frac{d}{d\alpha}\{[\tfrac{4}{3}(u^0)^2 - \tfrac{1}{3}]_f\rho R^4\} + \frac{\sigma_{ij}}{\dot\alpha}\frac{4}{3}\,{}_f\rho v_i v_j R^4 = \frac{\rho_\nu R^4[1+\langle V\rangle]}{\dot\alpha t_\nu} - \frac{rR^4{}_f\rho(\frac{4}{3}u_0{}^2 - \frac{1}{3})}{\dot\alpha t_e}. \quad (4.10)$$

Here we have simplified the equation by introducing the relation ${}_fp = {}_f\rho/3$.

The quantity t_e/r can be considered an "effective" collision time for neutrino energy density production from electrons. The last equation shows the energy density of the electrons being increased by the work done by the changing anisotropy of the metric (the $\sigma_{ij}v_iv_j$ term), an effect which is present even when t_e and t_ν are infinite. In each interval $d\alpha$ the energy density is further increased by

$$[\rho_\nu R^4(1+\langle V\rangle)/\dot\alpha t_\nu]d\alpha,$$

energy supplied by colliding neutrinos, and diminished by

$$[r\,{}_f\rho(\tfrac{4}{3}u_0{}^2 - \tfrac{1}{3})/\dot\alpha t_e]d\alpha$$

energy taken from the electron fluid to produce more neutrinos.

Just as we have seen that the $T^{0\alpha}{}_{;\alpha} = 0$ conservation law yields information about ${}_fT^{00}$, we will use the $T^{i\alpha}{}_{;\alpha} = 0$ conservation laws to obtain information about ${}_fT^{0i}$.

This Bianchi identity can be written, for type IX homogeneous models,[1,16]

$$(R^3T_{0i})^\cdot = R^3\epsilon_{ijk}T_{jl}g^{kl}.$$

Hence

$$[R^3({}_f\rho + {}_fp)u_0u_i]^\cdot - [\textstyle\int W_\nu(\alpha,\alpha')d\alpha'\int d^3q\, q_if_{\alpha'}]^\cdot = R^3\epsilon_{ijk}({}_f\rho + {}_fp)u_ju_lR^{-2}e^{-2\beta}_{lk} - R^3\epsilon_{ijk}R^2e^{\beta}_{jp}e^{\beta}_{ls}\rho_\nu\left\langle\frac{\partial V}{\partial\beta_{ps}}\right\rangle R^{-2}e^{-2\beta}_{kl}. \quad (4.11)$$

We now impose the requirement that both the electron gas and the neutrinos have momentum directed in one of the principal directions of the matrix β_{ij}; this is a *nontumbling* situation; we shall take only T_{03} nonvanishing and $T_{03}(\nu)$ nonvanishing. In that case one also has $e^{-2\beta}_{3i} = 0$ for $i \neq 3$, and we see that the term on the right in Eq. (4.11) involving the fluid quantities vanishes.

We subsequently restrict ourselves to these nontumbling models. We find, for the second term on the left in Eq. (4.11),

$$-\frac{d}{dt}[\textstyle\int W_\nu(\alpha,\alpha')d\alpha'\int d^3q\, q_if_{\alpha'}] = \int\frac{W_\nu}{t_\nu}a_\nu(R'T')^4l_i(\alpha')\,d\alpha' - \int\frac{\delta(\alpha-\alpha')}{t_e}\int d^3q\, q_if_{\alpha'}d\alpha' - \int W_\nu(\alpha,\alpha')\,d\alpha'\int d^3q\,\dot q_if_{\alpha'}. \quad (4.12)$$

In the first term on the right, we have used Eq. (2.10). The last term on the right in Eq. (4.12) is

$$\int W_\nu(\alpha,\alpha')d\alpha'\int\epsilon_{ijk}\frac{q_jq^k}{q^0}f_{\alpha'}\,d^3q = R^3T_{jl}(\nu)e^{-2\beta_{kl}}\epsilon_{ijk}R^{-2},$$

[Since the geodesic equation in these type IX models is $q_i = -\epsilon_{ijk}q_jq^k/q^0$ (Ref. 19)]. This term will cancel the neutrino term on the right in Eq. (4.11), in view of Eq. (2.11). The first term on the right in Eq. (4.12) can clearly be written as $\rho_\nu R^4\langle l_i\rangle/t_\nu$. The remaining term represents the quantity $-R^3T_{0i}/t_e$ for a distribution of neutrinos which is comoving with and has the same temperature as the electrons at the instant α.

Hence (4.11) reads

$$[R^3({}_f\rho + {}_fp)u_0u_i]^\cdot + \rho_\nu R^4\langle l_i\rangle t_\nu^{-1} + (R^3/t_e)\,r({}_f\rho + {}_fp)u_0u_i = 0 \quad (4.13)$$

for this nontumbling case where only l_3 and u_3 are nonzero.

Equations (4.10) and (4.13) can in principle be solved to give information directly about ${}_f\rho$ and u_i. The expressions are very involved, however, and we will not compute them at this time.

5. THE PRODUCTION OF ENTROPY

Because there are supposedly dissipative interactions between the electron–photon fluid and the partly collisionless neutrinos, we expect a production of entropy, i.e., we expect

$$s^\mu{}_{;\mu} \equiv (s_{(e)}u^\mu)_{;\mu} + s_{(\nu)}{}^\mu{}_{;\mu} \geq 0,$$

where $s_{(e)}$ is the specific entropy density of the electron–photon fluid, and $s_{(\nu)}{}^\mu$ is the entropy flux associated with the neutrinos. The definitions of $s_{(e)}$ and u^α follow directly from thermodynamics and hydrodynamics. The quantities for the neutrinos will be obtained from the usual statistical mechanics formulas. The derivation in this section is a generalization of that of Marle[20] who considered multicomponent systems only briefly.

Now the rest frame fluid entropy density $s_{(e)}$ is given by

$$s_{(e)} = \tfrac{4}{3}a_eT^3, \quad (5.1)$$

where

$${}_f\rho = a_eT^4.$$

We shall proceed first to compute the divergence of the electron–photon fluid's entropy vector. In the basis $\{dt, \sigma^i\}$, the connection forms obey $\Gamma^\alpha_{\alpha j} = 0$, and hence we have simply (since we deal only with homogeneous quantities)

$$(s_{(e)}u^\mu)_{;\mu} \equiv R^{-3}(R^3s_{(e)}u^0)_{,0}. \quad (5.2)$$

Because of Eq. (5.1), we see that $s_{(e)}$ is proportional to ${}_f\rho^{3/4}$. We can, therefore, use the previous conservation laws, Eqs. (4.10) and (4.13), to compute the quantity (5.2). We use the identity

$$\ln[{}_f\rho R^3u_3u^0] = \ln[{}_f\rho R^4u^{0\,4/3}] + \ln[u_3/(Ru^{0\,1/3})]. \quad (5.3)$$

Recall that u_3 is the only nonvanishing component of u_i. Equation (4.13) gives the time derivative of the right side of Eq. (5.3). Hence we can compute $(R^3s_{(e)}u^0)^\cdot$, using Eq. (4.13), and with additional terms arising from the last term in Eq. (5.3). In order to reduce the resulting expression, we need to evaluate the quantity $(u_0^2)^\cdot$ which appears:

$$(u_0{}^2)^{\cdot} = (1 + R^{-2}e^{-2\beta}_{ij}u_iu_j)^{\cdot}$$
$$= -2\dot{\alpha}(u_0{}^2 - 1) + 2(e^{-2\beta}_{ij})^{\cdot}\, R^{-2}u_iu_j + R^{-2}e^{-2\beta}_{33}u_3u_3\frac{(u_3{}^2)^{\cdot}}{(u_3{}^2)}$$

where we inserted the nontumbling condition. Hence

$$(u_0{}^2)^{\cdot} = -2\dot{\alpha}(u_0{}^2 - 1) - 2\sigma_{ij}v^iv^j + 2(u_0{}^2 - 1)(\ln u_3)^{\cdot}. \tag{5.4}$$

Inserting the expression Eq. (5.4) and multiplying through by $3u_0{}^2$, we obtain

$$3u_0{}^2\frac{d}{dt}\ln[{}_f\rho R^4u_0{}^{4/3}] + r\frac{3u_0{}^2}{t_e} + 3u_0{}^2\frac{\frac{3}{4}\rho_\nu R^4\langle l_3\rangle/t_\nu}{{}_f\rho R^3u_3u_0} - \dot{\alpha}(2u_0{}^2 + 1) + (\ln u_3)^{\cdot}(2u_0{}^2 + 1) = -\sigma_{ij}v_iv_j. \tag{5.5}$$

Starting again from Eq. (4.10), we can again compute $(d/dt)\ln({}_f\rho R^4u_0{}^{4/3})$, obtaining

$$3u_0{}^2\frac{(u_0{}^2 - \frac{1}{4})}{(u_0{}^2 - 1)}\frac{d}{dt}\ln({}_f\rho R^4u_0{}^{4/3}) + (2u_0{}^2 + 1)\times[-\dot{\alpha} + (\ln u_3)^{\cdot}] + \frac{3u_0{}^2}{(u_0{}^2 - 1)}\frac{(u_0{}^2 - \frac{1}{4})}{t_e}r - \frac{\frac{3}{4}3u_0{}^2}{(u_0{}^2 - 1)}\frac{\rho_\nu R^4[1 + \langle V\rangle]}{t_\nu{}_f\rho R^4} = -\sigma_{ij}v_iv_j. \tag{5.6}$$

Subtracting Eq. (5.5) from Eq. (5.6) and multiplying by $(u_0{}^2 - 1)/3u_0{}^2$, we have

$$\frac{d}{dt}\ln(s_{(e)}u^0R^3) = \frac{3}{4}\frac{d}{dt}\ln({}_f\rho R^4u_0{}^{4/3}) = -\frac{3}{4}\frac{r}{t_e} + \frac{3}{4}\frac{\rho_\nu R^4[1 + \langle V\rangle]}{t_\nu{}_f\rho R^4} + (u_0{}^2 - 1)\frac{\frac{3}{4}\rho_\nu R^4\langle l_3\rangle/t_\nu}{{}_f\rho R^3u_3u_0}. \tag{5.7}$$

This is the equation giving the evolution of the entropy density of the photon–electron fluid.

It is useful to rewrite this expression using the relation $s_{(e)} = \frac{4}{3}{}_f\rho T^{-1}$ [Eq. (5.1)]:

$$(s_{(e)}u^0R^3)^{\cdot} = (1/RT)\,[-r\,{}_f\rho u^0R^4t_e^{-1} + u^0\rho_\nu R^4 \times (1 + \langle V\rangle)\,t_\nu^{-1} - (u_0{}^2 - 1)(\rho_\nu R^4)\,\langle l_3\rangle\, t_\nu^{-1}Ru_3^{-1}]. \tag{5.8}$$

The first term on the right represents loss of fluid entropy as electrons are destroyed to produce neutrinos, while the second term shows the increase in electron-photon entropy when energetic neutrinos collide to produce neutrinos. The final term can have either sign and represents an interaction depending on the matter flow in the two species.

We now turn to the neutrino entropy flux. We define it in the usual way[21,22] expressed in the $\{dt, \sigma^i\}$ basis

$$s^\mu{}_{(\nu)} = -R^{-3}\int q^\mu F\,\ln(F/c')\,\frac{d^3q}{q^0} \tag{5.9}$$

(for Maxwell–Boltzmann neutrinos). Hence,

$$s^0{}_{(\nu)} = -R^{-3}\int F\,\ln(F/c')\,d^3q, \tag{5.10}$$

where c' is a constant with the dimensions of F.

Now

$$s^\mu{}_{(\nu);\mu} = R^{-3}(R^3s^0{}_{(\nu)})^{\cdot},$$
$$(R^3s^0{}_{(\nu)})^{\cdot} = -\int \dot{F}[\ln(F/c') + 1]\,d^3q. \tag{5.11}$$

We use Eqs. (3.3) and (3.4) and rearrange terms slightly to write

$$(R^3s^0{}_{(\mu)})^{\cdot} = \int\left[\frac{F}{t_\nu} - \frac{F_t}{t_e}\right]\left[\ln\left(\frac{Ft_\nu^{-1}}{F_tt_e^{-1}}\right) + \ln\left(\frac{F_t}{c'}\right) + 1 + \ln(t_\nu t_e^{-1})\right]d^3q. \tag{5.12}$$

As we desired, the term $(Ft_\nu^{-1} - F_tt_e^{-1})\ln(Ft_\nu^{-1}F_t^{-1}t_e)$ in Eq. (5.12) is nonnegative. Let us consider the quantity $\ln(F_t/c')$. F_t is a neutrino distribution which is thermal at the current instant and has the same temperature and 4-velocity as the electron–photon fluid. Since we restrict the discussion to Maxwell–Boltzmann distributions, the logarithm is especially easy to compute:

$$\ln(F_t/c) = c'' + (q_\alpha u^\alpha/T), \tag{5.13}$$

with u^α the 4-velocity of the electron fluid. In the $\{dt, \omega^i\}$ orthonormal frame, we have

$$q_\alpha u^\alpha = -\overset{0}{q}{}^0u^0 + \overset{0}{q}_i v^i, \tag{5.14}$$

where we make use of $\overset{0}{u}{}^i = v^i$. In terms of the integration variable q_i, we have

$$q_\alpha u^\alpha = -R^{-1}(e^{-2\beta}_{ij}q_iq_j)^{1/2}u^0 + R^{-1}e^{-\beta}_{3k}q_kv^3, \tag{5.15}$$

recalling that we deal with a nontumbling situation so u^i has only a 3-component.

The constants $1 + \ln(t_\nu/t_e)$ appearing in Eq. (5.12) and the constant c'' from Eq. (5.13) implied in (5.12) contribute to what might be called the "rest entropy" part of the entropy change since they give the same change in entropy any time a neutrino is created or destroyed. We consequently assume that these constants have been adjusted to make the rest entropy per particle some convenient quantity; we shall set it equal to zero.

In computing the remaining neutrino entropy divergence terms, then, we need only consider those arising from Eq. (5.15). The terms involving F/t_ν are particularly easy to compute; use Eqs. (2.9), (2.10), and (4.6) to obtain

$$\int F\,q_\alpha u^\alpha d^3q = -u^0(R^4\rho_\nu)[1 + \langle V\rangle]R^{-1} + v^3e^{-\beta}_{3k}R^{-1}(\rho_\nu R^4)\langle l_k\rangle. \tag{5.16}$$

The calculations involving F_t are also straightforward because F_t is thermal with the same 4-velocity and temperature as the electron–photon fluid:

$$\int F_tR^{-1}(e^{-2\beta}_{ij}q_iq_j)^{1/2}d^3q = rR^3[({}_f\rho + {}_fp)u_0{}^2 - {}_fp] \tag{5.17a}$$

and

$$\int F_tR^{-1}e^{-\beta}_{3i}q_id^3q = rR^3({}_f\rho + {}_fp)u^0v^i. \tag{5.17b}$$

Hence

$$(R^3s^0{}_{(\nu)})^{\cdot} = (1/RT)[-u^0(\rho_\nu R^4)(1 + \langle V\rangle)t_\nu^{-1} + v^3e^{-\beta}_{3k}(\rho_\nu R^4)\langle l_k\rangle t_\nu^{-1} + r({}_f\rho R^4)u^0t_e^{-1}] + (\text{positive terms}) + (\text{"rest entropy" terms}). \tag{5.18}$$

The bracketed terms written out in Eq. (5.18) are precisely the negative of the fluid entropy flux divergence, Eq. (5.8). The only term for which this is not entirely obvious is the $\langle l_k \rangle$ term. But, in Eq. (5.8), we can write

$$-(u_0{}^2 - 1)\, Ru_3^{-1} = (R^{-2} e_{33}^{-2\beta} u_3 u_3) R u_3^{-1} = -e_{33}^{-2\beta} u_3 R^{-1} = -e_{33}^{-\beta} v_3.$$

The result then follows: This 2-fluid model which expresses collisional processes in terms of two distinct collision times predicts entropy increase in every situation.

6. DISCUSSION

Finding that the entropy increases in the model discussed here, a model supposed to describe dissipation in homogeneous cosmologies, is satisfying from a physical viewpoint. However, the result says very little specifically about questions currently of interest in such cosmologies, for instance, whether the dissipation acts to reduce the over-all anisotropy of the system. Intuitively we feel that an entropy producing interaction should provide a "frictional force" which would tend to bring the rest frame defined (in some way) by the neutrino distribution into coincidence with the fluid rest frame. That idea has not yet been completely analyzed, but we saw in Eq. (4.12) that there is an additional "torque" proportional to $T_{jl}{}^{(\nu)} e_{kl}^{-2\beta} \epsilon_{ijk}$ which tends to change the quantity $T_{0i}(\nu)$ and so no alignment of the rest frames seems possible so long as the anisotropy matrix β_{ij} is nonvanishing, except for special cases. This is presumably one mechanism for anisotropy dissipation, but a complete study of this mechanism is left for later investigation.

It should be pointed out that all results found here can be directly applied to the similar calculation in type I cosmologies simply by setting ϵ_{ijk} to zero wherever it appears. (The matter currents T_{0i} must also vanish to type I.) This is then a justification of the thermodynamics implicit in the approach to dissipation utilized in the discussion of dissipation in type I cosmological models.[6,7]

The present work can be completely divorced from the question of anisotropic cosmologies. The entire calculation becomes special relativistic simply by setting $R = 1$, $e_{ij}^{\beta} = \delta_{ij}$, setting ϵ_{ijk} to zero wherever it appears and setting $\rho_\nu[1 + \langle V \rangle] = T_{00}(\nu)$. The calculation then describes, via a two-time collision approximation, the interaction between a partly collisional neutrino gas and an electron–photon fluid streaming through it.

For instance, one may consider in special relativity the relative streaming of an electron–photon fluid with temperature T and 4-velocity u^α through a partly collisional neutrino gas which instantaneously has a thermal distribution with 4-velocity $\bar{u}^\alpha$ and temperature $\bar{T}$. Then

$$s^\mu{}_{(\nu);\mu} = -\frac{\rho_\nu}{\bar{T}} \frac{\bar{u}^0}{t_\nu} + \frac{r}{\bar{T} t_e} \{[({}_f\rho + {}_fp)\, u_0{}^2 - {}_fp]\, \bar{u}^0 - ({}_f\rho + {}_fp)\, u^0 u_3 \bar{u}_3\},$$

while

$$s^\mu{}_{(e);\mu} = -(r/Tt_e)\, {}_f\rho u^0 + (1/Tt_\nu) \{[(\rho_\nu + p_\nu)\bar{u}_0{}^2 - p_\nu]\, u^0 - (\rho_\nu + p_\nu)\, \bar{u}^0 u_3 \bar{u}_3\}.$$

The sum of these two,

$$s^\mu{}_{;\mu} = (\rho_\nu/t_\nu)\,[-\bar{u}^0/\bar{T} + (1/T)\,(\tfrac{4}{3}\bar{u}_0{}^2 u^0 - \tfrac{1}{3}u^0 - \tfrac{4}{3}\bar{u}^0 u_3 \bar{u}_3)] + ({}_f\rho r/t_e)\,[-u^0/T + (1/\bar{T}) \times (\tfrac{4}{3}u_0{}^2 \bar{u}^0 - \tfrac{1}{3}\bar{u}^0 - \tfrac{4}{3}u^0 u_3 \bar{u}_3)] \tag{6.1}$$

(where we used $p_\nu = \frac{1}{3}\rho_\nu$, ${}_fp = \frac{1}{3}\,{}_f\rho$) reduces to a more familiar form

$$s^\mu{}_{;\mu} = u^0 \left[\frac{\rho_\nu}{t_\nu}\left(\frac{1}{T} - \frac{1}{\bar{T}}\right) + \frac{{}_f\rho}{t_e}\left(\frac{1}{\bar{T}} - \frac{1}{T}\right)\right] \tag{6.2}$$

when the fluids are relatively at rest ($u_3 = \bar{u}_3$). Equation (6.2) does not appear to be relativisticly invariant becuase of the factor u^0. But t_e and t_ν are defined in the observer defined rest frame [see Eqs. (3.3) and (3.4)], and u^0 is simply the Lorentz factor to correct to fluid's rest frame. Hence Eq. (6.2) is in fact Lorentz invariant.[23] Equation (6.2), together with the fact that $s^\mu{}_{;\mu} \geq 0$ as shown earlier, may be interpreted to mean "the net heat flow is from the warmer to the cooler object."

---

* Supported in part by the National Science Foundation (Grant No. 20033).

1 R. A. Matzner, Ann. Phys. (N.Y.) **65**, 438 (1971).
2 R. A. Matzner, Ann. Phys. (N.Y.) **65**, 482 (1971).
3 A. H. Taub, Ann. Math. **53**, 472 (1951).
4 H. P. Robertson, Phil. Mag. **5**, 835 (1928).
5 A. G. Walker, Proc. London Math. Soc., Ser. 2, **42**, 90 (1936).
6 R. A. Matzner and C. W. Misner, Astrophys. J. **171**, 415 (1972).
7 R. A. Matzner, Astrophys. J. **171**, 433 (1972).
8 C. W. Misner, Astrophys. J. **151**, 431 (1968).
9 J. M. Stewart, Monthly Notices Roy. Astron. Soc. **145**, 347 (1969).
10 J. M. Stewart and C. B. Collins, Monthly Notices Roy. Astron. Soc. **153**, 419 (1971).
11 J. M. Stewart, "Relativistic Thermodynamics and Kinetic Theory with Applications to Cosmology," Proceedings of Summer School on Relativity and Cosmology, Cargèse, Corsica 1971 (to be published, 1972).
12 M. P. Ryan, Ph.D. thesis (University of Maryland, 1970).
13 R. A. Matzner, L. C. Shepley, and J. B. Warren, Ann. Phys. (N.Y.) **57**, 401 (1970). (Referred to as "MSW" in text).
14 C. W. Misner, Phys. Rev. Lett. **22**, 1071 (1969).
15 M. P. Ryan, Ann. Phys. (N.Y.) **65**, 506 (1971); **68**, 541 (1971).
16 R. A. Matzner, J. Math. Phys. **11**, 2432 (1970).
17 P. Jordan, J. Ehlers, and W. Kundt, Akad. Wiss Mainz **2**, 23 (1960).
18 J. Ehlers and R. K. Sachs, "Kinetic Theory and Cosmology", in *Astrophysics and General Relativity*, edited by M. Chrétien, S. Deser, and J. Goldstein (Gordon and Breach, London, 1971).
19 S. W. Hawking, Monthly Notices Roy. Astron. Soc. **142**, 129 (1969).
20 C. Marle, Ann. Inst. H. Poincaré **10**, 67 (1969).
21 W. Israel, J. Math. Phys. **4**, 1163 (1963).
22 J. L. Synge, *The Relativistic Gas* (North-Holland, Amsterdam, 1957).
23 C. Matzner, unpublished notes (1972).

Poincaré and TCP Invariance in the Determination of Wave Equations for Particles of Arbitrary Spin

M. Seetharaman and P. M. Mathews
Department of Theoretical Physics, University of Madras, Madras-25, India.
(Received 13 January 1971; Revised Manuscript Received 11 March 1971)

A brief outline is given of different types of approaches to the derivation of relativistic wave equations for free particles of arbitrary mass m and spin s. It is pointed out that all such equations can be subsumed in a standard form involving wavefunctions transforming according to the representation $D(0,s) \oplus D(s,0)$ of the Lorentz group. Generalizing earlier work employing this standard form, we determine all such equations (in a standard form) which are invariant under the Poincaré group and the *combined* operation $\Theta \equiv TCP$. It is shown that the resulting equations are automatically invariant under T, C, and P *separately*.

I. INTRODUCTION

Relativistic wave equations for the description of particles of arbitrary spin s and mass m have been under investigation for over four decades now, in fact ever since Dirac's discovery[1] of the celebrated spin $-\frac{1}{2}$ relativistic wave equation. The requirement of manifest covariance, which was such a powerful force in determining the form of the Dirac equation, continued for many years to play a dominant role in investigations on the higher-spin cases, and in fact little else was demanded of relativistic wave equations. By 1950 a number of alternative manifestly covariant forms had been discovered.[2] There seems to have been no systematic attempt during this period to formulate clearly the minimum requirements on relativistic wave equations for given spin and mass and to determine the class of all wave equations consistent with the requirements, or on the other hand to explore in detail the interrelation between the known equations. The outstanding work of Harish-Chandra[3] forms an exception to this statement. By determining the condition on the matrices β_μ, $\mu = 0, 1, 2, 3$, in order that the equation

$$(\beta_\mu \partial^\mu + im)\psi = 0, \tag{1}$$

should describe a particle of unique mass and spin, he has effectively provided a systematization[4] of all *manifestly covariant equations linear in the differential operator* ∂_μ. However, it is quite possible for an equation to be relativistically invariant without being manifestly so; nor is linearity in ∂_μ (or even locality) an essential part of relativistic invariance. The starting point of Harish-Chandra's investigation must therefore be considered to be not completely general. It was Foldy,[5] who for the first time set up relativistic wave equations abandoning *manifest* covariance, who emphasized that all that is needed is that solutions of the wave equation should provide a representation space for the Poincaré group. He observed that by choosing the wavefunction ψ to be made up of two $(2s+1)$—component parts transforming according to $(2s+1)$—dimensional Wigner irreducible representations[6] with the time translation generator given, respectively, by $\pm E = \pm(\mathbf{p}^2 + m^2)^{1/2}$, one gets almost trivially the relativistically invariant equation

$$i\frac{\partial\psi}{\partial t} = \rho_3 E\psi, \tag{2}$$

where ρ_3 is diagonal, being $+1$ on the upper half of the wavefunction and -1 on the lower half. A new point which emerged from his analysis was the apparent existence of a considerable amount of freedom in choosing the operators representing the effect of the discrete operations T, C, P on the wavefunction. In subsequent years, the fact that it is the irreducible representations of the Poincaré group (and not the homogeneous Lorentz group), which are closely associated with the concept of an elementary particle, has tended to get increasing emphasis. At the same time the structure of locally covariant[7] fields as expressed in terms of entities transforming according to the Wigner representations has become rather well understood and gained familiarity through the work of Stapp[8] (in the definition of M functions in S-matrix theory), Joos,[9] Weinberg,[10] Pursey,[11] and others. Fields transforming according to the $(0,s)$ and $(s,0)$ representations of the homogeneous Lorentz group have a special place in this context since they are in one-to-one correspondence with the Poincaré irreducible set of entities (eg., creation/annihilation operators) characterizing a particle of given spin s and mass m; and fields with any other transformation property (m,n) constructed for the same particle, being determined by the same Poincaré-invariant set, can be expressed in terms of the $(0,s)$ or $(s,0)$ types of fields. This fact is of some importance in that it enables *any* manifestly covariant equation to be reduced to a form which involves only these special types of fields.[12] In the spin-$\frac{3}{2}$ case, for instance, Shay, Song, and Good[13] have carried out the actual reduction of a variety of familiar manifestly covariant equations to the common standard form

$$i\frac{\partial\psi}{\partial t} = H\psi, \quad \psi = \begin{pmatrix}\psi(0,s)\\ \psi(s,0)\end{pmatrix}, \tag{3}$$

and shown that the Hamiltonian H obtained is the same in all cases, coinciding with what Weaver, Hammer, and Good[14] had constructed earlier by assuming the form (3). Thus the existence of a standard form like (3) makes it possible to decide whether various equations are really distinct or not. Consequently, the problem of determining all possible wave equations consistent with any desired invariance requirements becomes well posed. One merely has to determine all possible forms of H in (3).

Investigations with this end in view were initiated by one of us some time ago.[15] Equation (3) was used as the standard form and H was sought to be constructed in such a way that (3) would be invariant under the Poincaré group and under T, C, and P. Initially, the imposition of the boost-invariance condition was done only for boosts parallel to the momentum direction; it turned out, nevertheless, that if H is also required to have a unique limit as $\mathbf{p} \to 0$, there is only one possibility for H. However, this last condition proved too restrictive to permit quantization of the resulting equation for integer spin.[16] On relaxing it, Mathews and Ramakrishnan[17] found that there were four classes of possibilities for H; but it was realized later that this embarrassingly rich variety arose because invariance under transverse boosts had still not been insisted upon. This condition also has been taken into account in a recent paper,[18] and it has been proved that there are just *two* possible forms of H consistent with invariance of (3) under the Poincaré group and under T, C, and P.

The next question naturally was: "To what extent is the above result conditioned by the requirement of T, C, and P invariance? If this were relaxed, would there be more possibilities for H?" It is with this question that the present paper is concerned. A partial answer has been given already,[19] namely, that if Poincaré and space inversion invariance are demanded, then H must be *either* one of the two forms referred to above, in which case T and C invariance would also hold, *or else* it should have the TCP-violating form $H = E$. For the proof of this statement, only a slight generalization of the treatment of Ref. 18 is needed because the most general starting form of H is not altered greatly by dropping T and C invariance. But when P-invariance is also dropped, as we proceed to do now, it is more advantageous to choose a different way of presenting the most general form of H. This gives rise to differences in matters of detail between the treatment here and that in Ref. 18; but the general approach remains the same. We require Poincaré and TCP invariance here, but not T, C, and P separately. Under these conditions we show that once again the only solutions for H are those found in Ref. 18 with separate T, C, and P invariance. Thus, TCP invariance for *any* relativistic equation for free fields implies T, C, and P separately. This seems to be the first time that a result of such generality has been proved. Our result shows that the fact that particular equations like the Dirac equation possess invariance under the discrete transformations (though such invariance is not a requirement in their derivation), is not merely accidental but is part of a general picture.

In the next section, we present the derivation of the above result. For completeness all the basic steps will be given but calculational details will be kept down to a minimum since the treatment here is a straightforward generalization of earlier work.[18] Throughout this paper, ψ will be treated as a c-number field. We comment briefly on the quantizability and other aspects of our theory in the concluding section.

Before we proceed with the derivation, a few remarks on the definition of the T, C, P operators may be in order, especially in relation to the question which is sometimes raised as to the meaning of C in the c-number theory. It is well to remind oneself in this context that though the concept of charge conjugation finds its clearest expression in the q-number theory, it was defined in the first place for the c-number Dirac field as the transformation $\psi(\mathbf{x}, t) \to \kappa\psi^*(\mathbf{x}, t)$ (with κ a numerical metrix) which leaves the Dirac equation invariant except for changing the sign of the charge. It carries any solution with definite energy and momentum or angular momentum into another with the signs of these quantities reversed. It is well known that this transformation is necessarily *antilinear*. When ψ is a quantum field, it is possible to realize C by a linear (unitary) transformation in the Hilbert space, but it is well to remember that the effect of C on ψ is still the very same transformation $\psi \to \kappa\psi^*$ (with the asterisk interpreted now as Hermitian conjugation on the components of ψ). In the present work with c-number fields of arbitrary spin, C is defined, as in Refs. 5 and 15, as an antilinear operation which reverses the signs of the Poincaré generators. This is consistent with the case of the Dirac field referred to above, but unlike this case, the actual form of H in our Eq. (3) is not given in advance: We have to make a *simultaneous* determination, from considerations of invariance of the form (3) under the Poincaré group and under $\Theta \equiv TCP$, of the explicit forms of both H and C (as well as of a unitary P and an anti-unitary T which also are defined initially purely in terms of commutation relations as in Ref. 5). This is what we now proceed to do.

II. DETERMINATION OF THE INVARIANT WAVE EQUATIONS

A. The Invariance Conditions

If Eq. (3) is to be invariant under the Poincaré group, the Hamiltonian H must transform like $-P_0$ where $P_0 = -i\partial/\partial t$ is the time translation generator. This means that all commutation relations involving P_0 should continue to be valid when the time-differential operator $-P_0$ is replaced everywhere [including in the structure of the boost generator $\mathbf{K}$, see Eq. (8c)] by H which is to be constructed out of $\mathbf{p}$ and available matrices:

$$[H, P_\mu] = 0, \tag{4a}$$

$$[H, \mathbf{J}] = 0, \tag{4b}$$

$$[H, \mathbf{K}] = i\mathbf{P}. \tag{4c}$$

As regards $\Theta \equiv TCP$, its commutation relations are seen from Ref. 5 to be

$$\Theta P_0 = -P_0\Theta, \quad \Theta\mathbf{P} = -\mathbf{P}\Theta, \quad \Theta\mathbf{J} = \mathbf{J}\Theta, \quad \Theta\mathbf{K} = \mathbf{K}\Theta. \tag{5}$$

Since it takes $P_0 = -i\partial/\partial t$ to $i\partial/\partial t$, for Θ invariance of (3) we must have

$$\Theta H = -H\Theta. \tag{6}$$

For a unique mass m we need

$$H^2 = E^2 = (\mathbf{p}^2 + m^2). \tag{7}$$

Of course a unique spin is already ensured by the choice of the transformation property $(0, s) \oplus (s, 0)$ for ψ. In this representation, the Poincaré generators have the following forms:[15]

$$\mathbf{P} = \mathbf{p} \equiv -i\nabla, \tag{8a}$$

$$\mathbf{J} = \mathbf{x} \times \mathbf{p} + \mathbf{S}, \qquad \mathbf{S} = \begin{pmatrix} \mathbf{s} & 0 \\ 0 & \mathbf{s} \end{pmatrix}, \tag{8b}$$

$$\mathbf{K} = t\mathbf{p} + \mathbf{x}p_0 + i\lambda, \qquad \lambda = \begin{pmatrix} \mathbf{s} & 0 \\ 0 & -\mathbf{s} \end{pmatrix} \equiv \rho_3\mathbf{S}, \tag{8c}$$

where $\mathbf{s}$ are the familiar vector of $(2s + 1)$-dimensional angular momentum matrices, and ρ_1, ρ_2, and ρ_3 are Pauli matrices. We have to solve Eqs. (4), (6), and (7) for H, when $\mathbf{P}, \mathbf{J}$, and $\mathbf{K}$ are given by (8).

It is easy now to see that the most general translation and rotation invariant operator that can be constructed from the above quantities is some function of $p \equiv |\mathbf{p}|$, $\mathbf{S}\cdot\mathbf{p}$, ρ_1, ρ_2, and ρ_3, and it can evidently be written in partitioned form as a 2×2 matrix, the elements of which are arbitrary functions of p and of the $(2s + 1)$-dimensional matrix operator $(\mathbf{s}\cdot\mathbf{p})$. But since $(\mathbf{s}\cdot\mathbf{p})$ has just $(2s + 1)$ eigenvalues νp with $\nu = +s, +s-1, \ldots, -s+1, -s$, any function of it can be expressed as a sum of projection operators α_ν to the eigenvalues νp of $(\mathbf{s}\cdot\mathbf{p})$:

$$\alpha_\mu \alpha_\nu = \alpha_\mu \delta_{\mu\nu}, \qquad F(\mathbf{s}\cdot\mathbf{p}) = \sum_{\nu=-s}^{s} F(\nu p)\alpha_\nu. \tag{9}$$

Instead of α_ν one may use the related projection operators β_ν and γ_ν, $\nu \geqslant 0$, defined by[20]

$$\beta_\nu = \alpha_\nu + \alpha_{-\nu} \tag{10a}$$

and

$$\gamma_\nu = \alpha_\nu - \alpha_{-\nu}, \tag{10b}$$

$$\beta_\mu\beta_\nu = \gamma_\mu\gamma_\nu = \beta_\mu\delta_{\mu\nu}, \quad \beta_\mu\gamma_\nu = \gamma_\mu\delta_{\mu\nu}, \quad \nu \geqslant 0. \tag{10c}$$

These have the advantage that β_ν and γ_ν being even and odd, respectively, in $(\mathbf{s}\cdot\mathbf{p})$, have simple transformation properties under Θ:

$$\Theta\beta_\nu = \beta_\nu\Theta, \qquad \Theta\gamma_\nu = -\gamma_\nu\Theta. \tag{11}$$

Equation (11) follows from the commutation rules (5). Equation (5) also shows that the matrix η representing the effect of Θ as defined by

$$\Theta\psi(\mathbf{x}, t) = \eta\psi(-\mathbf{x}, -t) \tag{12}$$

must be of the form

$$\eta = e^{i\phi}\begin{pmatrix}1 & 0\\ 0 & \xi\end{pmatrix}, \qquad \xi = \pm 1. \tag{13}$$

The restriction to $\xi = \pm 1$ comes from the condition

$$\Theta^2 \sim 1. \tag{14}$$

The sign $\sim$ indicates equality to within a phase factor, which arises because multiplication of a wavefunction by a phase factor does not alter the physical state it represents. However, the actual value of this overall phase, shown as $e^{i\phi}$ in Eq. (13), is irrelavant for our purpose.

B. The form of H

Returning now to the form of H, it follows from what has been said above that on account of translation and rotation invariance, H must be expressible as

$$H = \sum_{\nu\geqslant 0}\begin{pmatrix} a_\nu\beta_\nu + b_\nu\gamma_\nu & c_\nu\beta_\nu + d_\nu\gamma_\nu\\ e_\nu\beta_\nu + f_\nu\gamma_\nu & g_\nu\beta_\nu + h_\nu\gamma_\nu\end{pmatrix}, \tag{15}$$

where the coefficients $a_\nu, b_\nu, \cdots$ are, in general, functions of p. It is useful to simplify the form (15) by taking account of Θ invariance before taking up boost invariance which leads to complicated equations. By applying (11), we find that in order to satisfy (6), the coefficients of β_ν and γ_ν in (15) must obey the relations

$$a_\nu = g_\nu = 0, \tag{16a}$$

$$(1+\xi)c_\nu = 0, \qquad (1+\xi)e_\nu = 0, \tag{16b}$$

$$(1-\xi)d_\nu = 0, \qquad (1-\xi)f_\nu = 0. \tag{16c}$$

It follows that if the special choice $\xi = -1$ is made in (13), the Hamiltonian becomes

$$H_1 = \sum_{\nu\geqslant 0}\begin{pmatrix} b_\nu\gamma_\nu & c_\nu\beta_\nu\\ e_\nu\beta_\nu & h_\nu\gamma_\nu\end{pmatrix}, \tag{17}$$

with

$$b_\nu^2 + c_\nu e_\nu = E^2, \qquad h_\nu = -b_\nu, \tag{18}$$

on account of the Klein–Gordon condition[21] evaluated with the help of (10c). If on the other hand one takes $\xi = +1$, then H becomes

$$H_2 = \sum_{\nu\geqslant 0}\begin{pmatrix} b_\nu\gamma_\nu & d_\nu\gamma_\nu\\ f_\nu\gamma_\nu & h_\nu\gamma_\nu\end{pmatrix}, \tag{19}$$

with

$$b_\nu^2 + d_\nu f_\nu = E^2, \qquad h_\nu = -b_\nu. \tag{20}$$

The one condition which still remains to be imposed on the above forms of H is that of boost invariance, Eq. (4c), which can be rewritten with the aid of the explicit form of $\mathbf{K}$, Eq. (8c), as

$$H(\nabla_p H) = [H, \boldsymbol{\lambda}] + \mathbf{p}. \tag{21}$$

Equation (21) imposes powerful constraints on the coefficients in H_1 and H_2. To exhibit these constraints by evaluating both sides of (21) with any of the forms (17) or (19) for H, one needs formulas for quantities like $\nabla_p[\sum_\nu F_\nu(p)\beta_\nu]$, $\nabla_p[\sum_\nu F_\nu(p)\gamma_\nu]$, $[\sum_\nu F_\nu(p)\beta_\nu, \mathbf{s}]$, and $[\sum_\nu F_\nu(p)\gamma_\nu, \mathbf{s}]$, where $F_\nu(p)$ are arbitrary functions of p. These can be easily deduced from the formulas we have derived in Ref. 18, and are given in the Appendix. They appear as linear combinations of the quantities $\beta_\nu\mathbf{s}$, $\gamma_\nu\mathbf{s}$, $\beta_\nu\boldsymbol{\tau}$, $\gamma_\nu\boldsymbol{\tau}$, $\beta_\nu\mathbf{p}$, and $\gamma_\nu\mathbf{p}$ which are linearly independent except for $\nu = s$ when they are related by Eqs. (A5). Here $\boldsymbol{\tau} = \mathbf{s}\times\mathbf{p}/p$. On substituting these formulas in (21) and equating coefficients of linearly independent quantities, we obtain a set of equations for $a_\nu, b_\nu, \ldots$, whose explicit forms depend on which of our two alternatives for the Hamiltonian is being considered. We now proceed to discuss separately the solution of the equations in the two cases, referring to the Appendix for the details of derivation of the equations.

1. *The Hamiltonian* H_1

On taking $H = H_1$, Eq. (17), with $h_\nu = -b_\nu$, Eq. (18), and evaluating the boost-invariance condition (21), one gets the following equations for all $\nu > \nu_0$, where ν_0 is the lowest nonnegative value of ν ($\nu_0 = 0$ or $\frac{1}{2}$ according as the spin is an integer or a half-integer):

$$b_\nu b_{\nu-1} + c_\nu e_{\nu-1} = E^2 - p(b_\nu - b_{\nu-1}), \tag{22a}$$

$$(p + b_{\nu-1})c_\nu = (b_\nu - p)c_{\nu-1}, \tag{22b}$$

$$(p + b_{\nu-1})e_\nu = (b_\nu - p)e_{\nu-1}, \tag{22c}$$

$$c_\nu e_{\nu-1} = e_\nu c_{\nu-1}, \tag{22d}$$

$$b_\nu\frac{db_\nu}{dp} + c_\nu\frac{de_\nu}{dp} = p, \tag{22e}$$

$$b_\nu\frac{db_\nu}{dp} + e_\nu\frac{dc_\nu}{dp} = p, \tag{22f}$$

$$b_\nu\frac{dc_\nu}{dp} - c_\nu\frac{db_\nu}{dp} = -2\nu c_\nu, \tag{22g}$$

$$b_\nu\frac{de_\nu}{dp} - e_\nu\frac{db_\nu}{dp} = -2\nu e_\nu. \tag{22h}$$

The solution of these coupled equations is quite straightforward. On multiplying (22b) by $e_{\nu-1}$ and

(22c) by $c_{\nu-1}$ and adding, one finds with the aid of (22d) that

$$(p + b_{\nu-1})c_\nu e_{\nu-1} = (b_\nu - p)c_{\nu-1}e_{\nu-1}. \tag{23}$$

Now replacing $c_{\nu-1}e_{\nu-1}$ in (23) by $(E^2 - b_{\nu-1}^2)$ from (18) and combining the resulting equation with (22a) to eliminate $c_\nu e_{\nu-1}$, one can deduce the following relation:

$$\frac{b_\nu}{E} = \frac{(b_{\nu-1}/E) + \tanh 2\theta}{1 + (b_{\nu-1}/E)\tanh 2\theta}, \qquad \nu = s, s-1, \ldots, \nu_0 + 1, \tag{24}$$

where

$$\cosh\theta = E/m \quad \text{and} \quad \sinh\theta = p/m. \tag{25}$$

Equation (24) immediately yields

$$b_\nu = E \tanh 2\nu\theta, \quad \nu = s, s-1, \ldots, \nu_0. \tag{26}$$

Another possible alternative $b_\nu = E \coth 2\nu\theta$ is ruled out when the case $\nu = \nu_0$ is considered. To see this, let us consider the half-integral spin case $\nu_0 = \frac{1}{2}$. (A similar procedure can be followed for integer spins, $\nu_0 = 0$). Equations (22) hold for $\nu = \nu_0 = \frac{1}{2}$ too, but as we have explained in a similar context in Ref. 18, the undefined symbols $b_{-1/2}, c_{-1/2}, e_{-1/2}$ which now appear are merely convenient notations for certain sums which occur in the derivation of the formulas (A6) and (A7). They turn out to be equal to $-b_{1/2}, c_{1/2}$, and $e_{1/2}$, respectively. With these replacements, Eq. (22a) for $\nu_0 = \frac{1}{2}$ becomes

$$-b_{1/2}^2 + c_{1/2}e_{1/2} = E^2 - 2pb_{1/2}, \tag{27}$$

which together with (18) shows that

$$b_{1/2} = E \tanh\theta. \tag{28}$$

This is inconsistent with $b_\nu = E \coth 2\nu\theta$, which is therefore ruled out. It is a relatively simple matter now to substitute (26) in Eqs. (22) and show that

$$c_\nu = e_\nu = E \operatorname{sech} 2\nu\theta. \tag{29}$$

Actually one could have $c_\nu = \alpha E \operatorname{sech} 2\nu\theta$ and $e_\nu = \alpha^{-1}E \operatorname{sech} 2\nu\theta$ instead of (29), α being an arbitrary constant independent of ν and p. But α can be reduced to unity by choosing a new wavefunction whose lower half is α times that of the original wavefunction. Since this change of scale of one irreducible part of the wavefunction does not affect the transformation properties, our entire procedure goes through unaltered, and so there is no loss of generality in taking $\alpha = 1$.

Thus we finally have

$$H_1 = \sum_{\nu \geq \nu_0} \begin{pmatrix} \gamma_\nu \tanh 2\nu\theta & \beta_\nu \operatorname{sech} 2\nu\theta \\ \beta_\nu \operatorname{sech} 2\nu\theta & -\gamma_\nu \tanh 2\nu\theta \end{pmatrix} \tag{30}$$

or in the notation used in Ref. 18,

$$H_1 = E\sum_{\nu\geq 0} \tanh 2\nu\theta \cdot C_\nu + \rho_1 E \sum_{\nu\geq 0} \operatorname{sech} 2\nu\theta \cdot B_\nu, \tag{31}$$

where the projection operators B_ν, C_ν used in the earlier paper are related to β_ν, γ_ν by

$$B_\nu = \begin{pmatrix} \beta_\nu & 0 \\ 0 & \beta_\nu \end{pmatrix}, \quad C_\nu = \begin{pmatrix} \gamma_\nu & 0 \\ 0 & -\gamma_\nu \end{pmatrix}. \tag{32}$$

2. *The Hamiltonian H_2*

In this case, the equations for $\nu > \frac{1}{2}$ can be shown to be identical to Eqs. (22) except for the replacements $c_\nu \to d_\nu$ and $e_\nu \to f_\nu$. However, the equation corresponding to $\nu_0 = \frac{1}{2}$ is

$$-b_{1/2}^2 - d_{1/2}f_{1/2} = E^2 - 2pb_{1/2}, \tag{33}$$

which does not follow from (27) by the above replacement. The reason is that here, $b_{-1/2} = -b_{1/2}$, $d_{-1/2} = -d_{1/2}$, and $f_{-1/2} = -f_{1/2}$, with minus signs in all these relations unlike in the case of $b_{-1/2}$, $c_{-1/2}$, and $e_{-1/2}$. The net result is that the solution for b_ν now is

$$b_\nu = E \coth 2\nu\theta \tag{34a}$$

and using this, one finds

$$-d_\nu = f_\nu = E \operatorname{cosech} 2\nu\theta. \tag{34b}$$

Thus we finally have

$$H_2 = E\sum_\nu \coth 2\nu\theta \cdot C_\nu + \rho_1 E \sum_\nu \operatorname{cosech} 2\nu\theta \cdot C_\nu, \tag{35}$$

with C_ν defined by (32).

III. DISCUSSION

The work of the last section shows that there are only two possible forms of H, given by (31) and (35), such that Eq. (3) is invariant under the Poincaré group and under $\Theta \equiv TCP$. These are precisely the forms which were obtained in Ref. 18 after imposing separate T, C, and P invariance. It follows therefore that quite generally, any relativistic wave equation for free fields of unique spin and nonvanishing mass which is invariant under TCP is automatically invariant under T, C, and P separately. Such a general result does not seem to have been known till now,[22] though of course there have been many discussions of the behavior of particular wave equations under the discrete transformations. As we have noted in the introduction, the use of a standard form is an essential prerequisite for a general treatment. Equation (3) is not manifestly covariant, nor is the Hamiltonian (31) or (35) local in configuration space, though the familiar manifestly covariant equations can be reduced to the form (3) with H given by (31) for half-integral spins and (35) for integral spins. It does not seem possible to prove within the ambit of c-number theory that (31) and (35) should apply, respectively, to the half-integral and integral spin cases. But such an association does become mandatory if ψ is made a q-number field. One of us has shown that this is demanded by the microcausality condition[23] even if one starts with the four classes of Hamiltonians found in Ref. 17 (where transverse boost invariance was not imposed). Taking this also into account, we find that the matrices representing the effect of the T, C, and P are uniquely determined [apart from arbitrary-phase factors like $e^{i\phi}$ in Eq. (13)] for any given spin. If

$$P\psi(\mathbf{x}, t) = \sigma\psi(-\mathbf{x}, t), \quad T\psi(\mathbf{x}, t) = \tau\psi^*(\mathbf{x}, t), \quad C\psi(\mathbf{x}, t) = \kappa\psi^*(\mathbf{x}, t), \tag{36}$$

one finds,[23] by referring to Cases I and III of Ref. 17 to which the Hamiltonians (31) and (35) belong, and using the explicit forms of σ, τ, κ for the corresponding cases as given in Ref. 15, that

$$\sigma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} e^{i\phi_P}, \quad \tau = \begin{pmatrix} \xi_s & 0 \\ 0 & \xi_s \end{pmatrix} e^{i\phi_T}, \quad \kappa = \begin{pmatrix} 0 & \xi_s \\ (-)^{2s}\xi_s & 0 \end{pmatrix} e^{i\phi_C}. \tag{37}$$

Here ξ_s is the $(2s+1)$-dimensional matrix defined by[24] $\xi_s \mathbf{s} \xi_s^{-1} = -\mathbf{s}^*$, and the arbitrary phase factors possible in σ, τ, and κ are explicitly indicated. Thus if the minimum requirements which we have made are granted, the kind of ambiguity in the choice of T, C, P operators which was discussed by Foldy[5] does not exist. Incidentally it may be observed as a check that the product of σ, τ, and κ taken in *any order* yields η, Eq. (13), to within a phase factor.

One final remark may be in order regarding a peculiarity of the Hamiltonian (35) which leads to a quantizable theory for integer spins. It does not tend to a unique limit as $\mathbf{p} \to 0$. This kind of difficulty for integer spins could be anticipated from the work of Pursey[11]; the source of the trouble has been traced[25] by an analysis of the Kemmer equation, to the fact that the first time derivative of the zero helicity part of ψ (unlike the projections to other helicities) is really undetermined. The fact that despite this "undesirable" aspect of the expression (35) no ambiguities or difficulties have arisen in our treatment here or in earlier work may be taken as justification for using the standard form (3) without reservations in all cases.

ACKNOWLEDGMENTS

One of us (M.S.) would like to thank J. Jayaraman for useful discussions and the Department of Atomic Energy for financial assistance in the form of a Junior Fellowship.

APPENDIX

To evaluate the two sides of Eq. (21) we need the following formulas:

$$\nabla_p\Big(\sum_\nu F_\nu(p)\beta_\nu\Big) = \frac{1}{2p}\sum_\nu\Big[(F_{\nu+1} - F_{\nu-1})\gamma_\nu \mathbf{s} - (F_{\nu+1} + F_{\nu-1} - 2F_\nu)\beta_\nu i\boldsymbol{\tau} - \Big(\nu(F_{\nu+1} - F_{\nu-1}) - 2p\frac{dF_\nu}{dp}\Big)\beta_\nu \frac{\mathbf{p}}{p}\Big]. \tag{A1}$$

$$\nabla_p\Big(\sum_\nu F'_\nu(p)\gamma_\nu\Big) = \frac{1}{2p}\sum_\nu\Big[(F'_{\nu+1} - F'_{\nu-1})\beta_\nu \mathbf{s} - (F'_{\nu+1} + F'_{\nu-1} - 2F'_\nu)\gamma_\nu i\boldsymbol{\tau} - \Big(\nu(F'_{\nu+1} - F'_{\nu-1}) - 2p\frac{dF'_\nu}{dp}\Big)\gamma_\nu \frac{\mathbf{p}}{p}\Big]. \tag{A2}$$

$$\Big[\sum_\nu F_\nu(p)\beta_\nu, \mathbf{s}\Big] = \frac{1}{2}\sum_\nu\Big[(-F_{\nu+1} - F_{\nu-1} + 2F_\nu)\beta_\nu \mathbf{s} + (F_{\nu+1} - F_{\nu-1})\gamma_\nu i\boldsymbol{\tau} + \nu(F_{\nu+1} + F_{\nu-1} - 2F_\nu)\gamma_\nu \frac{\mathbf{p}}{p}\Big]. \tag{A3}$$

$$\Big[\sum_\nu F'_\nu(p)\gamma_\nu, \mathbf{s}\Big] = \frac{1}{2}\sum_\nu\Big[(-F'_{\nu+1} - F'_{\nu-1} + 2F'_\nu)\gamma_\nu \mathbf{s} + (F'_{\nu+1} - F'_{\nu-1})\beta_\nu i\boldsymbol{\tau} + \nu(F'_{\nu+1} + F'_{\nu-1} - 2F'_\nu)\beta_\nu \frac{\mathbf{p}}{p}\Big] \tag{A4}$$

In these equations F_ν, F'_ν are arbitrary functions of p and $\boldsymbol{\tau} = \mathbf{s} \times \mathbf{p}/p$. The proof of these formulas would proceed on the same lines as that of similar results in Ref. 18. But since B_ν and C_ν in that paper are defined in terms of $\boldsymbol{\lambda}$ in exactly the same way as β_ν and γ_ν here in terms of $\mathbf{s}$, and since the commutation relations $[\lambda_i, \lambda_j] = i\epsilon_{ijk}\rho_3\lambda_k$ and $[s_i, s_j] = i\epsilon_{ijk}s_k$ (which are all that we use in the derivation) are identical except for the appearance of ρ_3, it is easy to convince oneself that the expressions above can be trivially obtained from Ref. 18 by simply making the replacements $\boldsymbol{\lambda} \to \mathbf{s}$, $B_\nu \to \beta_\nu$, and $C_\nu \to \gamma_\nu$, and dropping factors of ρ_3. In this fashion, one obtains (A1) and (A2) from Eq. (C2) of that paper and (A3) and (A4) from Eq. (C6) by observing that H in Eqs. (C2) and (C6) is nothing but the sum of $\sum_\nu c_\nu C_\nu$ and $\sigma\sum_\nu b'_\nu B_\nu$. Incidentally, one can deduce in a similar way from Ref. 18, Appendix B, that the quantities $\gamma_\nu \mathbf{s}$, $\beta_\nu i\boldsymbol{\tau}$, $\beta_\nu \mathbf{p}/p$, $\beta_\nu \mathbf{s}, \gamma_\nu i\boldsymbol{\tau}$, $\gamma_\nu \mathbf{p}/p$, which occur in Eqs. (A1)–(A4), are linearly independent for all $\nu \neq s$ and that for $\nu = s$,

$$\beta_s i\boldsymbol{\tau} + s\beta_s \mathbf{p}/p = \gamma_s \mathbf{s} \tag{A5a}$$

and

$$\gamma_s i\boldsymbol{\tau} + s\gamma_s \mathbf{p}/p = \beta_s \mathbf{s}. \tag{A5b}$$

The application of Eqs. (A1) to (A5) in Eq. (21) is straightforward though tedious. In the case of $H = H_1$, for instance, differentiating (17) (with $h_\nu = -b_\nu$), using (A1) and (A2), and then premultiplying by H_1 and simplifying with the help of (10c), one gets $H_1(\nabla_p H_1)$ in 2×2 partitioned form whose upper left-hand element is

$$\frac{1}{2p}\sum_\nu\Big[[b_\nu(b_{\nu+1} - b_{\nu-1}) + c_\nu(e_{\nu+1} - e_{\nu-1})]\gamma_\nu \mathbf{s} - [b_\nu(b_{\nu+1} + b_{\nu-1} - 2b_\nu) + c_\nu(e_{\nu+1} + e_{\nu-1} - 2e_\nu)]\beta_\nu i\boldsymbol{\tau} - \Big(\nu b_\nu(b_{\nu+1} - b_{\nu-1}) - 2pb_\nu\frac{db_\nu}{dp} + \nu c_\nu(e_{\nu+1} - e_{\nu-1}) - 2pc_\nu\frac{de_\nu}{dp}\Big)\beta_\nu \frac{\mathbf{p}}{p}\Big]. \tag{A6}$$

The element in the corresponding position for $[H_1, \boldsymbol{\lambda}] + \mathbf{p}$ found with the aid of (A3) and (A4), is

$$\tfrac{1}{2}\sum_\nu\{(2b_\nu - b_{\nu+1} - b_{\nu-1})\gamma_\nu \mathbf{s} + (b_{\nu+1} - b_{\nu-1})\beta_\nu i\boldsymbol{\tau} + [\nu(-2b_\nu + b_{\nu+1} + b_{\nu-1}) + 2p]\beta_\nu(\mathbf{p}/p)\}. \tag{A7}$$

On account of Eq. (21) the expressions (A6) and (A7) have to be equated. Actually the coefficients of $\gamma_\nu \mathbf{s}$, $\beta_\nu i\boldsymbol{\tau}$, and $\beta_\nu \mathbf{p}/p$ can be separately equated (for all $\nu \neq s$) on account of their linear independence. The three equations which are thus obtained (for each ν) and the nine other such equations which result from equating the remaining elements of $H_1(\nabla_p H_1)$ and $[H_1, \boldsymbol{\lambda}] + \mathbf{p}$ are not all independent. Picking out the independent ones and combining them suitably, one can eliminate $b_{\nu+1}$, $c_{\nu+1}$, and $e_{\nu+1}$. The set of equations (22) then results. It may be noted that the undefined quantities b_{s+1}, c_{s+1}, e_{s+1}, which appear in (A6) and (A7) from the terms with $\nu = s$, automatically drop out because of (A5), and the equations which then result for $\nu = s$ are identical to (22).

1 P. A. M. Dirac, Proc. Roy. Soc. (London) **117A**, 610 (1928).
2 For a survey of manifestly covariant wave equations, see for example: E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Hafner, New York, 1953); H. Umezawa, *Quantum Field Theory* (North Holland, Amsterdam, 1956); and Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, Oxford, London, 1969).
3 Harish-Chandra, Phys. Rev. **71**, 793 (1947); Proc. Roy. Soc. (London) **192A**, 195 (1947). See also H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945); *ibid.* **21**, 451 (1949); and also I. M. Gel'fand and A. M. Yaglom, Zh. Eksp. Teor. Fiz. **18**, 703 (1948) for a general analysis of linear wave equations which are however not constrained to yield unique mass and spin.
4 For a lucid presentation, with explicit evaluation of matrices obeying Harish-Chandra's conditions for various low values of spin, see A. Z. Capri, Phys. Rev. **178**, 2427 (1969).
5 L. L. Foldy, Phys. Rev. **102**, 568 (1956).
6 E. P. Wigner, Ann. Math. **40**, 149 (1939).
7 We call a field locally covariant if $\psi(x)$ and $\psi'(x')$ representing it *at one and the same physical point* as seen from two inertial reference frames are related by a purely numerical matrix. Foldy's wavefunction in Ref. 5 is not locally covariant.
8 H. P. Stapp, Phys. Rev. **125**, 2139 (1962). See also A. O. Barut, I. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963).
9 H. Joos, Fortschr. Physik **10**, 65 (1964).
10 S. Weinberg, Phys. Rev. **133B**, 1318 (1964).
11 D. L. Pursey, Ann. Phys. (N.Y.) **32**, 157 (1965).
12 In the integer-spin case the reduced equation involves singular operators. The source and precise nature of the singular behavior have been investigated in Ref. 25. The apparent lack of rigor in the use of such equations is outweighed by the advantage of having a common standard form for relativistic wave equations, and the fact that in any case the formal developments go through without ambiguities and lead to meaningful results. It is to be emphasized that it is not the attempt to set up wave *equations* in some particular standard form that is responsible for the emergence of unconventional features in the integer-spin case. Similar features are present, for instance, in the work of Weinberg (Ref. 10) where wave equations are of no concern: the *wavefunctions* which go into the construction of integer-spin fields of the $(0, s) \oplus (s, 0)$ type there have the peculiar property that for any momentum and *zero* helicity the positive and negative energy wavefunctions are *identical*.
13 D. Shay, H. S. Song, and R. H. Good, Jr., Nuovo Cimento Suppl. **3**, 455 (1965).
14 D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135B**, 241 (1964).
15 P. M. Mathews, Phys. Rev. **143**, 978 (1966).
16 P. M. Mathews, Phys. Rev. **155**, 1415 (1967).
17 P. M. Mathews and S. Ramakrishnan, Nuovo Cimento **50**, 339 (1967).
18 M. Seetharaman, J. Jayaraman, and P. M. Mathews, J. Math. Phys. **12**, 835 (1971).
19 M. Seetharaman, J. Jayaraman, and P. M. Mathews, J. Math. Phys. **12**, 1620 (1971)
20 It may be noted that the definition of $\alpha_\nu, \beta_\nu, \gamma_\nu$ are parallel to those of $\Lambda_\nu, B_\nu, C_\nu$ in Ref. 15, the only difference being that while the former are projection operators related to the $(2s+1)$-dimensional matrix $(\mathbf{s}\cdot\mathbf{p})$, the latter relate to the $2(2s+1)$-dimensional matrix $(\boldsymbol{\lambda}\cdot\mathbf{p})$. The two sets are connected by Eq. (32).
21 Besides (16) there is another possible solution of the Klein-Gordon condition, namely $E^2 = b_\nu^2 = h_\nu^2$, $e_\nu = c_\nu = 0$. At the present stage, there is nothing to forbid the choice of solutions of this type for some values of ν and Eq. (16) for other values of ν. But this would be inconsistent with Eq. (21) below, required for boost invariance.
22 R. M. Santilli and P. Roman, "*TCP* Invariance Implies Separate T, C, and P Invariance," Boston University (Preprint, 1969), have claimed that *TCP* implies T, C, and P quite generally, even in the presence of interactions. However their proof rests upon very strong assumptions (of analyticity of Wightman functions with respect to two independent complex 4-vectors) for which there seems to be no real justification.
23 P. M. Mathews, J. Math. Phys. Sci. (Madras) **1**, 197 (1967). See also P. M. Mathews in *Lectures in Theoretical Physics*, Boulder Summer Institute, 1969, (Gordon and Breach, New York, 1971).
24 U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959), Appendix C.
25 M. Seetharaman, J. Jayaraman, and P. M. Mathews, Nuclear Phys. **19B**, 625 (1970).

Bounds for Solution of Nonlinear Wave-Wave Interacting Systems with Well-Defined Phase Description

P. K. C. Wang
School of Engineering and Applied Science, University of California, Los Angeles, California 90024
(Received 19 January 1972)

In this paper, upper and lower bounds for the solutions of nonlinear three-wave interacting systems with well-defined phase description are established. The results are used to deduce a sufficient condition for the nonexistence of explosive instabilities.

1. INTRODUCTION

The existence of explosive instabilities in various nonlinear wave-wave interacting systems such as plasmas have been investigated both theoretically and experimentally in recent years.[1-6] In most theoretical studies, the random phase approach is used. Consequently, the results do not yield information regarding the phases of waves for instability. It has been pointed out that phase effects may lead to significant modifications in the dynamics of wave-wave interactions.[7] In a recent paper by Wilhelmsson, Stenflo, and Engelmann,[2] a necessary condition for explosive instability in a three-wave system with well-defined phase description is presented. In this paper, upper and lower bounds for the solutions of the system equations governing the wave-wave interactions are established. The results are used to deduce a sufficient condition for the nonexistence of explosive instabilities. Although the results are obtained only for the three-wave system, the same approach is applicable to systems involving any finite number of interacting waves.

2. BASIC EQUATIONS

In this work, the following set of complex ordinary differential equations governing a pure three-wave interaction is considered:

$$\frac{da_0}{dt} = j\omega_0 a_0 + \mu_0 a_1 a_2, \qquad \frac{da_1}{dt} = j\omega_1 a_1 + \mu_1 a_0 a_2^*, \qquad \frac{da_2}{dt} = j\omega_2 a_2 + \mu_2 a_0 a_1^*, \tag{1}$$

with initial data at $t = 0$,

$$a_i(0) = a_{i0}, \quad i = 0, 1, 2, \tag{2}$$

where $j = \sqrt{-1}$ and $(\cdot)^*$ denotes complex conjugate; a_i and ω_i correspond to the complex amplitude and frequency of the ith normal mode respectively and the complex parameters μ_i are the coupling coefficients. For the case of a magnetized plasma with particle drift motions, explicit expressions for μ_i are given by Wilhelmsson.[7] It is evident that $a_0 = a_1 = a_2 = 0 + j0$ satisfy the following equilibrium equations corresponding to (1):

1 P. A. M. Dirac, Proc. Roy. Soc. (London) **117A**, 610 (1928).
2 For a survey of manifestly covariant wave equations, see for example: E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Hafner, New York, 1953); H. Umezawa, *Quantum Field Theory* (North Holland, Amsterdam, 1956); and Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, Oxford, London, 1969).
3 Harish-Chandra, Phys. Rev. **71**, 793 (1947); Proc. Roy. Soc. (London) **192A**, 195 (1947). See also H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945); *ibid.* **21**, 451 (1949); and also I. M. Gel'fand and A. M. Yaglom, Zh. Eksp. Teor. Fiz. **18**, 703 (1948) for a general analysis of linear wave equations which are however not constrained to yield unique mass and spin.
4 For a lucid presentation, with explicit evaluation of matrices obeying Harish-Chandra's conditions for various low values of spin, see A. Z. Capri, Phys. Rev. **178**, 2427 (1969).
5 L. L. Foldy, Phys. Rev. **102**, 568 (1956).
6 E. P. Wigner, Ann. Math. **40**, 149 (1939).
7 We call a field locally covariant if $\psi(x)$ and $\psi'(x')$ representing it *at one and the same physical point* as seen from two inertial reference frames are related by a purely numerical matrix. Foldy's wavefunction in Ref. 5 is not locally covariant.
8 H. P. Stapp, Phys. Rev. **125**, 2139 (1962). See also A. O. Barut, I. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963).
9 H. Joos, Fortschr. Physik **10**, 65 (1964).
10 S. Weinberg, Phys. Rev. **133B**, 1318 (1964).
11 D. L. Pursey, Ann. Phys. (N.Y.) **32**, 157 (1965).
12 In the integer-spin case the reduced equation involves singular operators. The source and precise nature of the singular behavior have been investigated in Ref. 25. The apparent lack of rigor in the use of such equations is outweighed by the advantage of having a common standard form for relativistic wave equations, and the fact that in any case the formal developments go through without ambiguities and lead to meaningful results. It is to be emphasized that it is not the attempt to set up wave *equations* in some particular standard form that is responsible for the emergence of unconventional features in the integer-spin case. Similar features are present, for instance, in the work of Weinberg (Ref. 10) where wave equations are of no concern: the *wavefunctions* which go into the construction of integer-spin fields of the $(0, s) \oplus (s, 0)$ type there have the peculiar property that for any momentum and *zero* helicity the positive and negative energy wavefunctions are *identical*.
13 D. Shay, H. S. Song, and R. H. Good, Jr., Nuovo Cimento Suppl. **3**, 455 (1965).
14 D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135B**, 241 (1964).
15 P. M. Mathews, Phys. Rev. **143**, 978 (1966).
16 P. M. Mathews, Phys. Rev. **155**, 1415 (1967).
17 P. M. Mathews and S. Ramakrishnan, Nuovo Cimento **50**, 339 (1967).
18 M. Seetharaman, J. Jayaraman, and P. M. Mathews, J. Math. Phys. **12**, 835 (1971).
19 M. Seetharaman, J. Jayaraman, and P. M. Mathews, J. Math. Phys. **12**, 1620 (1971)
20 It may be noted that the definition of $\alpha_\nu, \beta_\nu, \gamma_\nu$ are parallel to those of $\Lambda_\nu, B_\nu, C_\nu$ in Ref. 15, the only difference being that while the former are projection operators related to the $(2s + 1)$-dimensional matrix $(\mathbf{s}\cdot\mathbf{p})$, the latter relate to the $2(2s + 1)$-dimensional matrix $(\boldsymbol{\lambda}\cdot\mathbf{p})$. The two sets are connected by Eq. (32).
21 Besides (16) there is another possible solution of the Klein-Gordon condition, namely $E^2 = b_\nu^2 = h_\nu^2$, $e_\nu = c_\nu = 0$. At the present stage, there is nothing to forbid the choice of solutions of this type for some values of ν and Eq. (16) for other values of ν. But this would be inconsistent with Eq. (21) below, required for boost invariance.
22 R. M. Santilli and P. Roman, "*TCP* Invariance Implies Separate T, C, and P Invariance," Boston University (Preprint, 1969), have claimed that *TCP* implies *T*, *C*, and *P* quite generally, even in the presence of interactions. However their proof rests upon very strong assumptions (of analyticity of Wightman functions with respect to two independent complex 4-vectors) for which there seems to be no real justification.
23 P. M. Mathews, J. Math. Phys. Sci. (Madras) **1**, 197 (1967). See also P. M. Mathews in *Lectures in Theoretical Physics*, Boulder Summer Institute, 1969, (Gordon and Breach, New York, 1971).
24 U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959), Appendix C.
25 M. Seetharaman, J. Jayaraman, and P. M. Mathews, Nuclear Phys. **19B**, 625 (1970).

Bounds for Solution of Nonlinear Wave-Wave Interacting Systems with Well-Defined Phase Description

P. K. C. Wang

School of Engineering and Applied Science, University of California, Los Angeles, California 90024

(Received 19 January 1972)

In this paper, upper and lower bounds for the solutions of nonlinear three-wave interacting systems with well-defined phase description are established. The results are used to deduce a sufficient condition for the nonexistence of explosive instabilities.

1. INTRODUCTION

The existence of explosive instabilities in various nonlinear wave-wave interacting systems such as plasmas have been investigated both theoretically and experimentally in recent years.[1-6] In most theoretical studies, the random phase approach is used. Consequently, the results do not yield information regarding the phases of waves for instability. It has been pointed out that phase effects may lead to significant modifications in the dynamics of wave-wave interactions.[7] In a recent paper by Wilhelmsson, Stenflo, and Engelmann,[2] a necessary condition for explosive instability in a three-wave system with well-defined phase description is presented. In this paper, upper and lower bounds for the solutions of the system equations governing the wave-wave interactions are established. The results are used to deduce a sufficient condition for the nonexistence of explosive instabilities. Although the results are obtained only for the three-wave system, the same approach is applicable to systems involving any finite number of interacting waves.

2. BASIC EQUATIONS

In this work, the following set of complex ordinary differential equations governing a pure three-wave interaction is considered:

$$\frac{da_0}{dt} = j\omega_0 a_0 + \mu_0 a_1 a_2, \quad \frac{da_1}{dt} = j\omega_1 a_1 + \mu_1 a_0 a_2^*, \quad \frac{da_2}{dt} = j\omega_2 a_2 + \mu_2 a_0 a_1^*, \tag{1}$$

with initial data at $t = 0$,

$$a_i(0) = a_{i0}, \quad i = 0, 1, 2, \tag{2}$$

where $j = \sqrt{-1}$ and $(\cdot)^*$ denotes complex conjugate; a_i and ω_i correspond to the complex amplitude and frequency of the ith normal mode respectively and the complex parameters μ_i are the coupling coefficients. For the case of a magnetized plasma with particle drift motions, explicit expressions for μ_i are given by Wilhelmsson.[7] It is evident that $a_0 = a_1 = a_2 = 0 + j0$ satisfy the following equilibrium equations corresponding to (1):

$$a_0 = j\mu_0 a_1 a_2/\omega_0,\ a_1 = j\mu_1 a_0 a_2^*/\omega_1,\ a_2 = j\mu_2 a_0 a_1^*/\omega_2. \tag{3}$$

To determine the existence of nontrivial equilibrium solutions (a_0^e, a_1^e, a_2^e), let

$$a_i^e = |a_i^e| \exp(j\theta_i), \quad \mu_i/\omega_i = \eta_i \exp(j\psi_i), \quad i = 0, 1, 2. \tag{4}$$

Elementary computations show that nontrivial equilibrium solutions exist if the quantities A_i defined by

$$A_0^2 \equiv \omega_1\omega_2^*/\mu_1\mu_2^*, \quad A_1^2 \equiv -\omega_0\omega_2/\mu_0\mu_2,$$
$$A_2^2 \equiv -\omega_0\omega_2/\mu_0\mu_1 \tag{5}$$

are real and positive; moreover, the ψ_i are related by

$$\psi_0 = (2n-1)\pi - \psi_1, \quad \psi_1 = 2m\pi + \psi_2,$$
$$m, n = 0, \pm 1, \pm 2, \cdots. \tag{6}$$

Assuming that (5) and (6) are satisfied, the equilibrium solutions are given by

$$a_i^e = A_i \exp(j\theta_i), \quad i = 0, 1, 2, \tag{7}$$

where $(\theta_0, \theta_1, \theta_2)$ is any ordered-triplet of real numbers satisfying the following phase consistency conditions:

$$\theta_0 - \theta_1 - \theta_2 = (2n - \tfrac{1}{2})\pi - \psi_1, \quad n = 0, \pm 1, \pm 2, \cdots. \tag{8}$$

It is evident that nontrivial equilibrium solutions exist only for special values of parameters μ_i and ω_i, $i = 0, 1, 2$. These solutions have time-invariant magnitudes and phase angles.

In the sequel, various upper and lower bounds for the norm of the solutions of (1) and (2) will be derived first. Then, a sufficient condition for the nonexistence of explosive instabilities will be established. Finally, special solutions for the system equations will be developed.

3. ESTIMATES OF SOLUTIONS

Let C_3 denote the normed complex vector space of ordered triplets of complex numbers $\mathbf{a} \equiv (a_0, a_1, a_2)$, whose norm is defined by

$$\|\mathbf{a}\| \equiv \left(\sum_{i=0}^{2} a_i a_i^*\right)^{1/2}. \tag{9}$$

First, certain differential inequalities involving the norm of the solutions of (1) will be developed.

Lemma 1: Let $\mathbf{a}(t) \equiv (a_0(t), a_1(t), a_2(t))$ be any solution of (1). Then, $\|\mathbf{a}(t)\|$ satisfies

$$-\alpha^-\|\mathbf{a}(t)\|^2 - \beta\|\mathbf{a}(t)\|^4 \le \frac{d\|\mathbf{a}(t)\|^2}{dt} \le \alpha^+\|\mathbf{a}(t)\|^2 + \beta\|\mathbf{a}(t)\|^4, \tag{10}$$

where

$$\alpha^\pm \equiv \min\{\max\{\lambda_{00}^\pm, \lambda_{12}^\pm, \lambda_{21}^\pm\}, \max\{\lambda_{01}^\pm, \lambda_{10}^\pm, \lambda_{22}^\pm\},$$
$$\times \max\{\lambda_{00}^\pm, \lambda_{01}^\pm, \lambda_{22}^\pm\}, \max\{\lambda_{01}^\pm, \lambda_{12}^\pm, \lambda_{21}^\pm\},$$
$$\times \max\{\lambda_{02}^\pm, \lambda_{10}^\pm, \lambda_{21}^\pm\}, \max\{\lambda_{02}^\pm, \lambda_{11}^\pm, \lambda_{20}^\pm\}\}, \tag{11a}$$

$$\beta \equiv \tfrac{1}{2} \max_i\{|\mu_i|\}, \tag{11b}$$

where

$$\lambda_{ij}^\pm \equiv |\mu_j| \mp 2\mathrm{Im}(\omega_i), \quad i, j = 0, 1, 2. \tag{12}$$

Proof: Consider the following equations obtained by multiplying the ith equation in (1) and its complex conjugate by a_i^* and a_i, respectively:

$$a_i^* \frac{da_i}{dt} = j\omega_i a_i^* a_i + \mu_i a_0^* a_1 a_2, \tag{13a}$$

$$a_i \frac{da_i^*}{dt} = -j\omega_i^* a_i^* a_i + \mu_i^* a_0 a_1^* a_2^*. \tag{13b}$$

Adding the above equations and summing over i lead to

$$\frac{d}{dt}\sum_{i=0}^{2} a_i a_i^* \equiv \frac{d}{dt}\|\mathbf{a}\|^2 = \sum_{i=0}^{2}[-2\mathrm{Im}(\omega_i)|a_i|^2 + \mu_i a_0^* a_1 a_2 + \mu_i^* a_0 a_1^* a_2^*]. \tag{14}$$

Thus,

$$\frac{d}{dt}\|\mathbf{a}\|^2 \le \sum_{i=0}^{2}[-2\mathrm{Im}(\omega_i)|a_i|^2 + 2|\mu_i|\ |a_0|\ |a_1|\ |a_2|], \tag{15a}$$

$$\frac{d}{dt}\|\mathbf{a}\|^2 \ge \sum_{i=0}^{2} -[2\mathrm{Im}(\omega_i)|a_i|^2 + 2|\mu_i|\ |a_0|\ |a_1|\ |a_2|]. \tag{15b}$$

From the elementary inequality $2xy \le x^2 + y^2$, it is evident that

$$2|\mu_i||a_0||a_1||a_2| \le \begin{cases} |\mu_i|(|a_0|^2 + |a_1|^2|a_2|^2), \\ |\mu_i|(|a_1|^2 + |a_0|^2|a_2|^2), \\ |\mu_i|(|a_2|^2 + |a_0|^2|a_1|^2). \end{cases} \tag{16}$$

For each i, the right-hand side of (16) is to be selected so that the coefficients α^+ and α^- in the desired estimates (10) are minimized. Evidently, if each inequality in (16) is used exactly once for $i = 0, 1, 2$, then there are only six possible sets of coefficients $\lambda_{ij}^\pm$ given in (11a). Consequently,

$$-h(a_0, a_1, a_2) - \alpha^-\|\mathbf{a}\|^2 \le \frac{d\|\mathbf{a}\|^2}{dt} \le \alpha^+\|\mathbf{a}\|^2 + h(a_0, a_1, a_2), \tag{17}$$

where α^+ and α^- given in (11) correspond to the smallest coefficients for the $\|\mathbf{a}\|^2$ terms in (17) and where

$$h(a_0, a_1, a_2) \equiv |\mu_0|\ |a_1|^2|a_2|^2 + |\mu_1|\ |a_0|^2|a_2|^2 + |\mu_2|\ |a_0|^2|a_1|^2. \tag{18}$$

Using the following algebraic inequality,

$$\begin{aligned} |a_1|^2|a_2|^2 + |a_0|^2|a_2|^2 + |a_0|^2|a_1|^2 &= \tfrac{1}{2}|a_2|^2(|a_0|^2 + |a_1|^2) \\ &+ \tfrac{1}{2}|a_1|^2(|a_0|^2 + |a_2|^2) \\ &+ \tfrac{1}{2}|a_0|^2(|a_1|^2 + |a_2|^2) \\ &\le \tfrac{1}{2}|a_2|^2\|\mathbf{a}\|^2 + \tfrac{1}{2}|a_1|^2\|\mathbf{a}\|^2 \\ &+ \tfrac{1}{2}|a_0|^2\|\mathbf{a}\|^2 = \tfrac{1}{2}\|\mathbf{a}\|^4, \end{aligned} \tag{19}$$

h is bounded by

$$h(a_0, a_1, a_2) \le \max_i\{|\mu_i|\}(|a_1|^2|a_2|^2 + |a_0|^2|a_2|^2 + |a_0|^2|a_1|^2) \le \beta\|\mathbf{a}\|^4. \tag{20}$$

The desired estimate (10) follows trivially from (17) and (19). Thus, the proof is complete.

To obtain upper and lower bounds for $\|\mathbf{a}(t)\|$ from (10), use will be made of the following result:

Lemma 2: Let w be a real-valued nonnegative differentiable function defined for $0 \leqslant t \leqslant T$ satisfying

$$-g^-(w(t)) \leqslant \frac{dw(t)}{dt} \leqslant g^+(w(t)), \qquad w(0) = w_0 \geqslant 0. \tag{21}$$

If the $g^{\pm}(u)$ are continuous and nondecreasing for $u \geqslant 0$ and $g^{\pm}(u) > 0$ for $u > 0$, $i = 1, 2$, then

$$w(t) \leqslant G_+^{-1}(t) \quad \text{for } 0 \leqslant t < t^+, \tag{22}$$

$$w(t) \geqslant G_-^{-1}(-t) \quad \text{for } 0 \leqslant t < t^-, \tag{23}$$

where

$$G_{\pm}(w) \equiv \int_{w_0}^{w} \frac{d\xi}{g^{\pm}(\xi)}, \qquad w \geqslant w_0 \geqslant 0; \tag{24}$$

and the interval $[0, t^{\pm})$ is a subset of the domain of definition of $G_{\pm}^{-1}$.

Proof: Consider the lower inequality in (21):

$$\frac{dw(t)}{dt} + g^-(w(t)) \geqslant 0, \qquad w(0) = w_0 \geqslant 0. \tag{25}$$

Since $g^-(w(t)) > 0$ for $w(t) > 0$, (25) can be rewritten as

$$0 \leqslant \frac{1}{g^-(w(t))} \frac{dw(t)}{dt} + 1 = \frac{d}{dt} G_-(w(t)) + 1. \tag{26}$$

Integrating (26) leads to

$$G_-(w(t)) - G_-(w_0) \geqslant -t. \tag{27}$$

Estimate (23) follows directly from the monotonicity of G_-^{-1} and $G_-(w_0) = 0$. Estimate (22) can be established in a similar manner.

Note that $G_+^{-1}(t)$ and $G_-^{-1}(-t)$ are simply the solutions of $dw/dt = g^+(w)$ and $dw/dt = -g^-(w)$, respectively, with initial condition $w(0) = w_0$. Applying Lemma 2 to (10) in Lemma 1 leads to the following estimates for $\|\mathbf{a}(t)\|^2$:

Theorem 1: Let $\mathbf{a}(t) \equiv (a_0(t), a_1(t), a_2(t))$ be a solution of (1) satisfying initial data (2). Assume that $\alpha^{\pm}$ and β defined in (11) are nonzero. Then:

(i) for $\alpha^- < 0$,

$$\|\mathbf{a}(t)\|^2 \leqslant q^+(t) \quad \text{for } 0 \leqslant t < t_1^+, \tag{28}$$

$$\|\mathbf{a}(t)\|^2 \geqslant p^-(t) \quad \text{for } t \geqslant 0 \text{ and } \|\mathbf{a}(0)\|^2 > |\alpha^-|/\beta; \tag{29}$$

(ii) for $\alpha^+ < 0$,

$$\|\mathbf{a}(t)\|^2 \leqslant p^+(t) \quad \text{for } 0 \leqslant t < t_2^+ \text{ and } \|\mathbf{a}(0)\|^2 > |\alpha^+|/\beta, \tag{30}$$

$$\|\mathbf{a}(t)\|^2 \geqslant q^-(t) \quad \text{for } t \geqslant 0, \tag{31}$$

where

$$q^{\pm}(t) \equiv \|\mathbf{a}(0)\|^2 \exp(\pm \alpha^{\pm} t)[1 + (\beta/\alpha^{\pm})\|\mathbf{a}(0)\|^2 \times (1 - \exp(\pm \alpha^{\pm} t)]^{-1}, \tag{32}$$

$$p^{\pm}(t) \equiv [\|\mathbf{a}(0)\|^2 + \beta/\alpha^{\pm}] \exp(\alpha^{\pm} t) \times (\{1 \pm [1 + (\beta/\alpha^{\pm})\|\mathbf{a}(0)\|^2]\}[1 - \exp(\alpha^{\pm} t)])^{-1} + |\alpha^{\pm}|/\beta, \tag{33}$$

$$\begin{aligned} t_1^+ &= (1/\alpha^+) \ln[1 + \alpha^+/\beta\|\mathbf{a}(0)\|^2], \\ t_2^+ &= (1/\alpha^+) \ln\{1 + \alpha^+/[\beta\|\mathbf{a}(0)\|^2 + \alpha^+]\}. \end{aligned} \tag{34}$$

Proof: First, consider case (i). Since $\alpha^- < 0$ implies $\lambda_{ij}^- < 0$ for all $i, j = 0, 1, 2$ and $\mathrm{Im}(\omega_i) > 0$ for all i. Consequently, $\alpha^+ > 0$. Let $w(t)$ in Lemma 1 be $\|\mathbf{a}(t)\|^2$. Thus,

$$\frac{dw(t)}{dt} \leqslant \alpha^+ w(t) + \beta w^2(t) \equiv g^+(w(t)) \tag{35}$$

with $w(0) = \|\mathbf{a}(0)\|^2$. Since $\beta > 0$, $g^+(w)$ increases monotonically from zero for $w \geqslant 0$. The upper bound

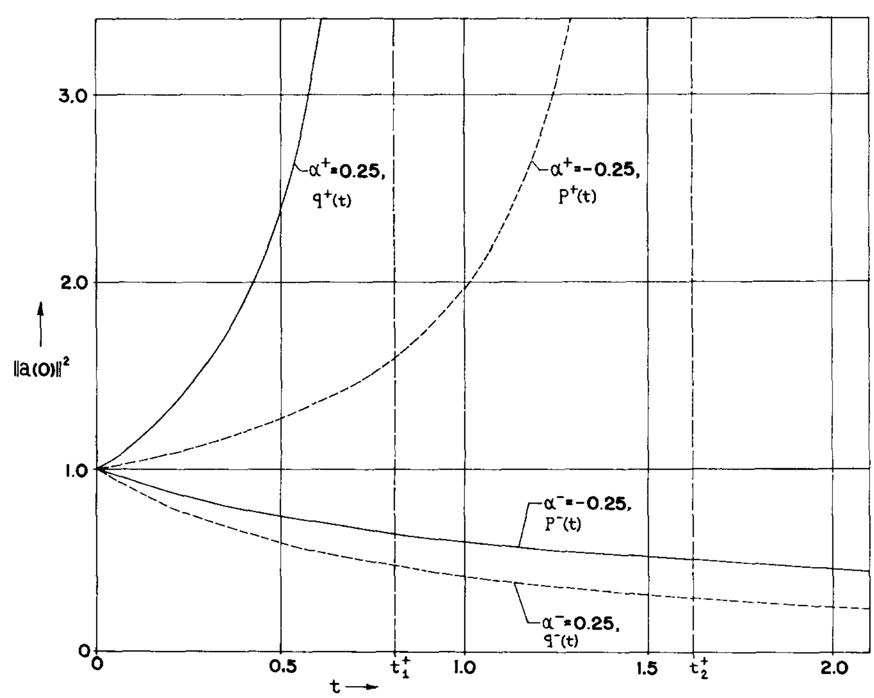

FIG. 1. Graphs of $q^{\pm}(t)$ and $p^{\pm}(t)$ for $\beta = 1.0$ and various values of α^+ and α^-.

(28) for $\|\mathbf{a}(t)\|^2$ can be obtained by direct application of Lemma 2 with

$$G_+(w(t)) = \int_{w(0)}^{w(t)} \frac{d\xi}{\xi(\alpha^+ + \beta\xi)} = \frac{1}{\alpha^+}\ln\left(\frac{w(t)[\alpha^+ + \beta w(0)]}{w(0)[\alpha^+ + \beta w(t)]}\right) \tag{36}$$

and

$$G_+^{-1}(t) = \frac{w(0)\exp(\alpha^+ t)}{1 + (\beta/\alpha^+)w(0)[1 - \exp(\alpha^+ t)]} \tag{37}$$

defined for $0 \leqslant t < t^+$.

To derive lower bound (29), let $w(t) \equiv \|\mathbf{a}(t)\|^2 + \alpha^-/\beta$. From the lower bound in (10),

$$\frac{dw(t)}{dt} = \frac{d\|\mathbf{a}(t)\|^2}{dt} \geqslant -\alpha^-\|\mathbf{a}(t)\|^2 - \beta\|\mathbf{a}(t)\|^4$$
$$= -\alpha^- w(t) - \beta w^2(t) \equiv -g^-(w(t)) \tag{38}$$

with $w(0) = \|\mathbf{a}(0)\|^2 + \alpha^-/\beta$. Clearly, $g^-(w)$ increases monotonically from zero for $w \geqslant 0$. Estimate (29) follows directly from (23) in Lemma 2. Estimates (30) and (31) for the case with $\alpha^+ < 0$ can be derived in a similar manner, since $\alpha^+ < 0$ implies $\mathrm{Im}(\omega_i) > 1$ for all i and $\alpha^- > 0$.

Note that the upper bound $q^+(t)$ in (28) tends to infinity as $t \to t_1^+$ and $p^+(t)$ in (30) tends to infinity as $t \to t_2^+$ if $\|\mathbf{a}(0)\|^2 \neq 2|\alpha^+|/\beta$. The lower bounds (29) and (31) tend to $|\alpha^-|/\beta$ and zero, respectively, as $t \to \infty$. Evidently, from (29), the condition $\alpha^- < 0$ or $\mathrm{Im}(\omega_i) < 0$ for all i implies the nonexistence of waves whose amplitudes tend to zero as $t \to \infty$. The time-domain behavior of the derived bounds is shown in Fig. 1.

4. EXPLOSIVE INSTABILITIES

By explosive instability, it is meant there exist solutions of (1) which become unbounded over some finite or infinite time interval. In the theory of ordinary differential equations, explosive instability corresponds to having solutions with *finite* or *infinite escape time*.[8,9] Obviously, the boundedness of the norm of solutions for all $t \geqslant 0$ implies nonexistence of explosive instabilities. On the other hand, the existence of a lower bound for the norm of solutions which becomes unbounded in finite time implies explosive instability. Note that the lower bounds (29) and (31) do not have this property for any α, since $\beta > 0$.

To derive a sufficient condition for the nonexistence of explosive instabilities, the upper bound in (10) is rewritten as follows:

$$\frac{d\|\mathbf{a}(t)\|^2}{dt} \leqslant -\|\mathbf{a}(t)\|^2[-\alpha^+ - \beta\|\mathbf{a}(t)\|^2]. \tag{39}$$

Clearly, if $\alpha^+ < 0$, then for any $\mathbf{a}(0)$ lies in the set Ω defined by

$$\Omega = \{\mathbf{a} \in C_3 : \|\mathbf{a}\| < (|\alpha^+|/\beta)^{1/2}\}, \tag{40}$$

its corresponding solution $\mathbf{a}(t)$ satisfies $\|\mathbf{a}(t)\| < (|\alpha^+|/\beta)^{1/2}$ for all $t > 0$, or Ω is an invariant set of system (1). Since Ω is bounded, there are no explosive solutions. This simple result can be stated as

Theorem 2: A sufficient condition for the nonexistence of explosive instabilities for system (1) with initial condition $\mathbf{a}(0)$ at $t = 0$ is that $\alpha^+ < 0$ and $\|\mathbf{a}(0)\| < (|\alpha^+|/\beta)^{1/2}$.

When explosive solutions exist, the estimate (28) or (30) in Theorem 1 provides an upper bound for the growth of solutions. Moreover, the expressions for t_i^+ given by (34) provide lower bounds for the escape times.

5. SPECIAL SOLUTIONS

In this section, the special cases for which closed-form solutions exist will be examined. The development begins with the derivation of equations which are equivalent to the original system (1).

Consider the derivative of the first equation in (1):

$$\frac{d^2a_0}{dt^2} = j\omega_0\frac{da_0}{dt} + \mu_0\frac{d(a_1a_2)}{dt}. \tag{41}$$

Multiplying the second and third equations in (1) by a_2 and a_1, respectively, and combining the results lead to

$$\frac{d}{dt}(a_1a_2) = j(\omega_1 + \omega_2)a_1a_2 + \mu_2a_0|a_1|^2 + \mu_1a_0|a_2|^2. \tag{42}$$

Substituting (42) into (40) and eliminating a_1a_2 by means of the first equation in (1) result in an alternate differential equation for a_0:

$$\frac{d^2a_0}{dt^2} - j(\omega_0 + \omega_1 + \omega_2)\frac{da_0}{dt} - \omega_0(\omega_1 + \omega_2)a_0$$
$$= \mu_0a_0(\mu_2|a_1|^2 + \mu_1|a_2|^2). \tag{43}$$

Similarly, the following differential equations for a_1 and a_2 can be derived:

$$\frac{d^2a_1}{dt^2} - j(\omega_0 + \omega_1 - \omega_2^*)\frac{da_1}{dt} - \omega_1(\omega_0 - \omega_2^*)a_1$$
$$= \mu_1a_1(\mu_2^*|a_0|^2 + \mu_1|a_2|^2), \tag{44}$$

$$\frac{d^2a_2}{dt^2} - j(\omega_0 - \omega_1^* + \omega_2)\frac{da_2}{dt} - \omega_2(\omega_0 - \omega_1^*)a_2$$
$$= \mu_2a_2(\mu_1^*|a_0|^2 + \mu_0|a_1|^2). \tag{45}$$

Also, by eliminating $a_0^*a_1a_2$ and $a_0a_1^*a_2^*$ among the equations in (14), the following relation for $|a_i(t)|^2$ can be readily established:

$$\mathrm{Im}(\mu_1\mu_2^*)\frac{d}{dt}|a_0|^2 - \mathrm{Im}(\mu_0\mu_2^*)\frac{d}{dt}|a_1|^2$$
$$+ \mathrm{Im}(\mu_0\mu_1^*)\frac{d}{dt}|a_2|^2 = -2\,\mathrm{Im}(\omega_0)\,\mathrm{Im}(\mu_1\mu_2^*)|a_0|^2$$
$$+ 2\,\mathrm{Im}(\omega_0)\,\mathrm{Im}(\mu_0\mu_2^*)|a_1|^2$$
$$- 2\,\mathrm{Im}(\omega_2)\,\mathrm{Im}(\mu_0\mu_1^*)|a_2|^2. \tag{46}$$

Moreover, for any pair of equations in (14), the condition

$$\mu_j\left(\frac{d}{dt}|a_i|^2 + 2\,\mathrm{Im}(\omega_i)|a_i|^2\right) = \mu_i\left(\frac{d}{dt}|a_j|^2 + 2\,\mathrm{Im}(\omega_j)|a_j|^2\right) \tag{47}$$

is satisfied for $i, j = 0, 1, 2$ and $i \neq j$.

Now, consider the special case where $a_0(0)$ is an arbitrary complex number; $a_1(0)$ and $a_2(0)$ are chosen such that their corresponding solutions satisfy

$$|a_1(t)|^2 = -\omega_0\omega_2/\mu_0\mu_2 \equiv A_1^2 \geqslant 0,$$
$$|a_2(t)|^2 = -\omega_0\omega_2/\mu_0\mu_1 \equiv A_2^2 \geqslant 0 \tag{48}$$

for all t. It can be readily verified that (48) satisfies the consistency condition (47) provided that $\omega_1^*\omega_2$ is real. To establish the existence of solutions $a_1(t)$ and $a_2(t)$ satisfying (48), let

$$a_i(t) = A_i \exp(j\phi_i t), \quad i = 1, 2, \tag{49}$$

where the ϕ_i are undetermined real numbers. Substituting (49) into the last two equations in (1) gives

$$\begin{aligned} jA_1(\phi_1 - \omega_1) \exp(j\phi_1 t) &= \mu_1 a_0 A_2 \exp(-j\phi_2 t), \\ jA_2(\phi_2 - \omega_2) \exp(j\phi_2 t) &= \mu_2 a_0 A_1 \exp(-j\phi_1 t). \end{aligned} \tag{50}$$

Eliminating a_0 among the above equations and making use of the explicit expressions for A_1 and A_2, one has the following consistency condition:

$$\omega_1\phi_2 = \omega_2\phi_1. \tag{51}$$

This condition implies that ω_1 must be a scalar multiple of ω_2. Thus, solutions in the form of (49) exist if ϕ_1 and ϕ_2 satisfy (51).

Now, an equation for $a_0(t)$ can be derived by substituting (48) into (43):

$$\frac{d^2a_0}{dt^2} - j(\omega_0 + \omega_1 + \omega_2)\frac{da_0}{dt} = 0, \tag{52}$$

whose solution with initial conditions $a_0(0)$ and $\dot{a}_0(0)$ has the form:

$$\begin{aligned} a_0(t) = a_0(0) &+ [\dot{a}_0(0)/j(\omega_0 + \omega_1 + \omega_2)] \\ &\times \exp[j(\omega_0 + \omega_1 + \omega_2)t], \quad t \geqslant 0. \end{aligned} \tag{53}$$

From (46), one also has a differential equation for $|a_0(t)|^2$:

$$\frac{d}{dt}|a_0|^2 + 2\,\mathrm{Im}(\omega_0)|a_0|^2 = \Delta_0, \tag{54}$$

where

$$\begin{aligned} \Delta_0 \equiv [2\omega_0/\mu_0\,\mathrm{Im}(\mu_1\mu_2^*)][(\omega_1/\mu_1)\,\mathrm{Im}(\omega_2)\,\mathrm{Im}(\mu_0\mu_1^*) \\ - (\omega_2/\mu_2)\,\mathrm{Im}(\omega_0)\,\mathrm{Im}(\mu_0\mu_2^*)]. \end{aligned} \tag{55}$$

The solution for (54) with initial condition $|a_0(0)|^2$ at $t = 0$ has the form

$$\begin{aligned} |a_0(t)|^2 = [|a_0(0)|^2 - \Delta_0/2\,\mathrm{Im}(\omega_0)] \exp[-2\,\mathrm{Im}(\omega_0)t] \\ + \Delta_0/2\,\mathrm{Im}(\omega_0) \end{aligned} \tag{56}$$

defined on an interval $0 \leqslant t \leqslant t_s$ for which $|a_0(t)|^2$ is nonnegative. Note that for $\Delta_0 < 0$ and $\mathrm{Im}(\omega_0) > 0$, the right-hand side of (56) becomes negative for sufficiently large t and it equals to zero when $t = t_s$ defined by

$$t_s = [1/2\,\mathrm{Im}(\omega_0)] \ln [1 - 2\,\mathrm{Im}(\omega_0)|a_0(0)|^2/\Delta_0]. \tag{57}$$

This implies that $|a_0(t)|^2$ goes to zero in *finite* time t_s or the wave amplitude of the zeroth mode decays to zero at $t = t_s$. This is also true for the case where $\Delta_0 < 0$, $\mathrm{Im}(\omega_0) < 0$, and $|a_0(0)|^2 < \Delta_0/2\,\mathrm{Im}(\omega_0)$. For the critical case where $|a_0(0)|^2 = \Delta_0/2\,\mathrm{Im}(\omega_0)$, $|a_0(t)|^2 = \Delta_0/2\,\mathrm{Im}(\omega_0)$ for all $t \geqslant 0$. When $\mathrm{Im}(\omega_0) < 0$, the solution $|a_0(t)|^2 \to \infty$ as $t \to \infty$ when $\Delta_0 > 0$ for all $|a_0(0)|^2 > 0$ or when $\Delta_0 < 0$ with $|a_0(0)|^2 > \Delta_0/2\,\mathrm{Im}(\omega_0)$. Finally, when $|a_0(0)|^2 = 0$, its corresponding solution $a_0(t)$ has the following properties:

(i) $|a_0(t)|^2 \to \infty$ as $t \to \infty$ if $\Delta_0 > 0$ and $\mathrm{Im}(\omega_0) < 0$;

(ii) $|a_0(t)|^2 \to \Delta_0/2\,\mathrm{Im}(\omega_0)$ as $t \to \infty$ if $\Delta_0 > 0$ and $\mathrm{Im}(\omega_0) > 0$.

There is no solution when $\Delta_0 < 0$, $\mathrm{Im}(\omega_0) > 0$, and $|a_0(0)|^2 = 0$.

Similar solutions are obtainable for $a_1(t)$ when

$$\begin{aligned} |a_0(t)|^2 &= \omega_1\omega_2^*/\mu_1\mu_2^* \equiv A_0^2 \geqslant 0 \quad \text{and} \\ |a_2(t)|^2 &= -\omega_0\omega_2/\mu_0\mu_1 \equiv A_2^2 \geqslant 0 \quad \text{for all } t, \end{aligned}$$

and for $a_2(t)$ when

$$\begin{aligned} |a_0(t)|^2 &= \omega_1\omega_2^*/\mu_1\mu_2^* \equiv A_0^2 \geqslant 0 \quad \text{and} \\ |a_1(t)|^2 &= -\omega_0\omega_2/\mu_0\mu_2 \equiv A_1^2 \geqslant 0 \quad \text{for all } t. \end{aligned}$$

6. CONCLUDING REMARKS

It has been shown that estimates for the solutions of a three-wave system with well-defined phase description are readily obtainable. These estimates provide qualitative information on the solution behavior from which a sufficient condition for the nonexistence of explosive instabilities can be deduced. It is not difficult to obtain similar results for systems involving any finite number of interacting waves using the approach presented here. The application of these results to specific systems arising in physical situations such as in plasmas will be discussed elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by a U.K. National Research Council Senior Post-Doctoral Fellowship at the University of Warwick, Coventry, and in part by AFOSR Grants No. AFOSR-68-1547 and 72-2303.

1 V. M. Dikasov, L. I. Rudakov, and D. D. Ryutov, Zh. Eksp. Teor. Fiz. 48, 913 (1965) [Sov. Phys. JETP **21**, 608 (1965)].
2 H. Wilhelmsson, L. Stenflo, and F. Engelmann, J. Math. Phys. **11**, 1738 (1970).
3 T. Sato, Phys. Fluids **14**, 2426 (1971).
4 R. E. Aamodt and M. Sloan, Phys. Fluids **11**, 2218 (1968).
5 P. Martin and B. D. Fried, Bull. Amer. Phys. Soc. **15**, 1443 (1970).
6 F. Engelmann and H. Wilhelmsson, Z. Naturforsch. **24a**, 206 (1969).
7 H. Wilhelmsson, J. Plasma Phys. **3**, 215 (1969).
8 L. Markus, Rend. Sem. Mat. Politecnico Torino **11**, 271 (1952).
9 A. Wintner, Amer. J. Math. **67**, 277 (1945).

Some Comments on the Behavior of Acceleration Waves of Arbitrary Shape

Ray M. Bowen
Rice University, Houston, Texas

and

Peter J. Chen
Sandia Laboratories, Albuquerque, New Mexico

(Received 9 December 1971; Revised Manuscript Received 16 February 1972)

This work concerns an application of the general results of Bailey and Chen to the study of acceleration waves of arbitrary shape. A modification of one of their main theorems is also presented; it applies to converging waves. For both converging and diverging waves, general expressions are given in terms of the initial principal curvatures. For diverging waves, a necessary and sufficient condition is given for the initial critical amplitude to be nonzero. In addition, Bailey and Chen's asymptotic formulas are specialized so as to apply to acceleration waves of arbitrary shape.

1. INTRODUCTION

In two recent papers Bailey and Chen[1,2] considered the local and global behavior of the solutions of the Bernoulli equation[3]

$$\frac{da}{dt} = -\mu(t)a + \beta(t)a^2 \tag{1.1}$$

subject to the following conditions:

(C1) μ and β are defined and integrable on every finite subinterval of $[0, \infty)$;

(C2) β is of fixed sign on $[0, \infty)$;

(C3) $\liminf_{t\to\infty} |\beta(t)| \neq 0$.

In Ref. 2, (C3) was replaced by the weaker condition

(C4) $\int_0^\infty |\beta(t)|\,dt = \infty$.

The motivation for the study of (1.1) is that it arises in a wide class of problems involving acceleration waves in materials. Bailey and Chen[4] showed, in particular, that when (a), (b), and (c) or (d) are satisfied and $\operatorname{sgn} a(0) = \operatorname{sgn}\beta(t)$, and

(i) if $|a(0)| > \alpha$, then

$$\lim_{t\to t_\infty} |a(t)| = \infty; \tag{1.2}$$

(ii) if $|a(0)| < \alpha$, then

$$\liminf_{t\to\infty} |a(t)| = 0, \tag{1.3}$$

where α is the *critical initial amplitude* defined by[5]

$$\alpha = \frac{1}{\int_0^\infty |\beta(t)|\, e^{-\int_0^t \mu(\tau)d\tau} dt}, \tag{1.4}$$

and t_∞ is a finite time defined by[6]

$$\int_0^{t_\infty} \beta(t) e^{-\int_0^t \mu(\tau)d\tau} dt = \frac{1}{a(0)}. \tag{1.5}$$

The concept of the critical amplitude was first introduced by Coleman and Gurtin[7] for the case of *one-dimensional* acceleration waves with μ and β being *constants*. The result (1.5) was first given by Coleman, Greenberg and Gurtin.[8]

For the class of problems involving diverging waves of arbitrary shape, the conditions (C1)–(C4) hold and thus Bailey and Chen's result described above can be applied. Not all problems involving curved acceleration waves of arbitrary shape satisfy (C1)–(C4). However, by altering Bailey and Chen's conditions, one can prove a theorem which, along with Bailey and Chen's, enables one to prove certain general results about diverging and converging waves of arbitrary shape.

2. APPLICATION TO WAVES OF ARBITRARY SHAPE

Given the differential equation (1.1), we shall adopt the following assumptions about the functions μ and β:

(A1) There exists a finite positive number t^* such that μ and β are defined and integrable on every subset of $[0, t^*)$;

(A2) $\lim_{t\to t^*} \int_0^t \mu(\tau)d\tau = -\infty$;

(A3) β is of fixed sign on $[0, t^*)$.

Given (A1)–(A3) we shall prove the following theorem.

Theorem 1: Consider the differential equation (1.1). Let (A1)–(A3) hold, let $\operatorname{sgn} a(0) = \operatorname{sgn}\beta(t)$, and let

$$\gamma = 1\Big/\left[\int_0^{t^*} |\beta(t)| \exp\left(-\int_0^t \mu(\tau)d\tau\right)dt\right] \tag{2.1}$$

(i) If $|a(0)| > \gamma$, then there exists a time $\hat{t} > 0$ such that $\hat{t} < t^*$,

$$\int_0^{\hat{t}} \beta(t) \exp\left(-\int_0^t \mu(\tau)d\tau\right)dt = \frac{1}{a(0)} \tag{2.2}$$

and

$$\lim_{t\to\hat{t}} |a(t)| = \infty.$$

(ii) If $|a(0)| < \gamma$, then

$$\lim_{t\to t^*} |a(t)| = \infty.$$

(iii) If $|a(0)| = \gamma$, then $\hat{t} = t^*$ and conversely. In addition

$$\lim_{t\to t^*} |a(t)| = \infty.$$

Proof: The solution of (1.1) on $[0, t^*)$ is[9]

$$a(t) = a(0)\Big/\exp\left(\int_0^t \mu(\tau)\,d\tau\right) \times\left[1 - a(0)\int_0^t \beta(\tau)\exp\left(-\int_0^\tau \mu(s)ds\right)d\tau\right]. \tag{2.3}$$

(i) Since $a(0)$ and $\beta(t)$ are of the same sign, the second factor in the denominator of (2.3) is a continuous, monotone decreasing function of t, and has the value $1 - |a(0)|/\gamma$ at $t = t^*$. Thus, if $|a(0)| > \gamma$ then there is a unique time $\hat{t} < t^*$ such that (2.2) is satisfied and the second factor in the denominator of (2.3) vanishes. Obviously, $\lim_{t\to\hat{t}}|a(t)| = \infty$.

(ii) In order that $|a(0)| < \gamma$ be possible, it must be that $\gamma > 0$, thus the second factor in the denominator of (2.3) approaches a finite positive limit as $t \to t^*$. Because of (A2), the first factor in the denominator vanishes and, thus, $\lim_{t\to t^*}|a(t)| = \infty$.

(iii) If $|a(0)| = \gamma$, then (2.1) and (2.2) yield $\hat{t} = t^*$. Clearly, if $\hat{t} = t^*$, then these same equations yield $|a(0)| = \gamma$. Again, it is obvious that $\lim_{t\to t^*}|a(0)| = \infty$. ■

Henceforth, we shall also refer to γ, defined by (2.1), as critical initial amplitude.

Our next objective is to show how Bailey and Chen's theorem,[10] which we discussed in Sec. 1, and the above theorem can be applied to the study of waves of arbitrary shape. Our discussion will be limited to a special class of waves. Namely, to the class for which the waves propagate as families of *parallel surfaces* with constant normal speeds. This physical circumstance arises in a number of interesting examples. For example, the papers by Thomas,[11] Varley and Dunwoody,[12] Varley and Cumberbatch,[13] Varley,[14] Chen,[15,16,17] and Doria and Bowen[18] have this feature. The impact of this assumption is that $\mu(t)$ takes the form

$$\mu(t) = \mu_0 - \tfrac{1}{2}u_n\overline{K}(t), \tag{2.4}$$

where μ_0 is a constant, u_n is the constant normal speed (taken to be positive), and $\overline{K}(t)$ is the mean curvature. In addition to (2.4), $\beta(t)$ is a *nonzero* constant which we shall denote by β_0. For a parallel family of surfaces propagating with constant velocity the mean curvature has the representation[19]

$$\overline{K}(t) = (\overline{K}_0 - 2K_0u_nt)/(1 - \overline{K}_0u_nt + K_0u_n^2t^2), \tag{2.5}$$

where $\overline{K}_0 = \kappa_1 + \kappa_2$ is the initial mean curvature and $K_0 = \kappa_1\kappa_2$ is the initial total curvature; κ_1 and κ_2 being the initial principal curvatures. The assumption that $\beta(t) = \beta_0$ insures that (C2)–(C4), and (A3) are satisfied. Similarly, the statements involving β in (C1) and (A1) are satisfied. The validity of the remaining portions of (C1) and/or (A1) and (A2) require the examination of (2.4) and (2.5). It can be shown from these equations that

$$\int_0^t \mu(\tau)d\tau = \mu_0t + \ln|(1 - \kappa_1u_n t)^{1/2}|\,|(1 - \kappa_2u_nt)^{1/2}|. \tag{2.6}$$

Thus, if κ_1 and κ_2 obey the conditions

$$\kappa_1 \le 0, \quad \kappa_2 \le 0, \tag{2.7}$$

then (C1) is satisfied. Similarly, if one or both of the principal curvatures are positive then (A1) and (A2) are satisfied with t^* being the smallest positive root of

$$(1 - \kappa_1u_nt^*)(1 - \kappa_2u_nt^*) = 0. \tag{2.8}$$

Thus, we must consider two distinct cases. For Case I, we shall assume κ_1 and κ_2 obey (2.7); for Case II, we shall assume at least one of the principal curvatures is positive. Case I corresponds to diverging waves, while Case II corresponds to converging waves.

Case I: Diverging Waves

Bailey and Chen's theorem shows that when sgn$a(0)$ = sgnβ_0, (1.2) and (1.3) are valid, where now α and t_∞ are given by

$$\alpha = \left(|\beta_0|\int_0^\infty \frac{e^{-\mu_0t}}{(1 - \kappa_1u_nt)^{1/2}(1 - \kappa_2u_nt)^{1/2}}dt\right)^{-1} \tag{2.9}$$

and

$$\int_0^{t_\infty} \frac{e^{-\mu_0t}}{(1 - \kappa_1u_nt)^{1/2}(1 - \kappa_2u_nt)^{1/2}}\,dt = \frac{1}{\beta_0a(0)}. \tag{2.10}$$

When $|a(0)| > \alpha$, the finite time t_∞ is the time required for a shock wave to form. Equations (2.9) and (2.10) generalize a number of interesting results given in the references cited earlier. Equation (2.9) contains the interesting result that *for diverging parallel waves of arbitrary shape a necessary and sufficient condition for the existence of a nonzero critical initial amplitude is*

$$\mu_0 > 0. \tag{2.11}$$

In other words, when $\mu_0 \le 0$, the critical initial amplitude vanishes.

When (2.11) holds, it follows from (2.9) that

$$\frac{\partial\alpha}{\partial\mu_0} > 0. \tag{2.12}$$

Equations (2.11) and (2.12) show that *the critical initial amplitude α increases as μ_0 increases*. A similar analysis shows

$$\frac{\partial\alpha}{\partial|\kappa_\alpha|} > 0, \quad \kappa_\alpha \ne 0, \quad \alpha = 1, 2. \tag{2.13}$$

These results show that *as the magnitude of the initial curvatures is increased, the critical initial amplitude also increases*. One can also show from (2.10) that t_∞, the time required for the shock to form, is a strictly increasing function of μ_0.

Bailey and Chen[20] presented several asymptotic results which are interesting to specialize here. From their formula (7) we can write, whenever $|a(0)| > \alpha$,

$$a(t) = [1/\beta_0(t - t_\infty)][1 + o(1)] \tag{2.14}$$

as $t \to t_\infty$. Equation (2.14) gives the amplitude near the time t_∞. It is interesting to observe that the curvature only enters this expression through the time t_∞. Bailey and Chen's formula (8) tells us that whenever $|a(0)| > \alpha$,

$$t_\infty = [1/a(0)\beta_0][1 + o(1)] \tag{2.15}$$

as $|a(0)| \to \infty$. Thus for very large initial amplitudes, t_∞ is *independent* of the curvature.

An asymptotic expression which does depend upon the curvature arises if we assume $|a(0)| < \alpha$. In this case Bailey and Chen's equation (9) yields

$$a(t) = \frac{a(0)e^{-\mu_0t}}{(1 - \kappa_1\mu_nt)^{1/2}(1 - \kappa_2u_nt)^{1/2}}\cdot\frac{1}{(1 - |a(0)|/\alpha)} \times[1 + o(1)] \tag{2.16}$$

as $a(0) \to 0$. Bailey and Chen's (9) also holds in the limit as $t \to \infty$. In this limit (2.16) yields

$$a(t) = \frac{a(0)e^{-\mu_0t}}{u_n(\kappa_1\kappa_2)^{1/2}t}\cdot\frac{1}{(1 - |a(0)|/\alpha)}[1 + o(1)] \tag{2.17}$$

as $t \to \infty$ whenever κ_1 and κ_2 are not zero. In applying (2.16) and (2.17) it is important to note that the requirement $|a(0)| < \alpha$ requires α to be nonzero. Thus, μ_0 must obey the condition (2.11).

Case II: Converging Waves

Theorem 1 shows that when $\operatorname{sgn} a(0) = \operatorname{sgn}\beta_0$ and $|a(0)| > \gamma$, then $\lim_{t\to\hat{t}} |a(t)| = \infty$, where, by (2.1) and (2.2), γ and $\hat{t}$ are given by

$$\gamma = \left(|\beta_0| \int_0^{t^*} \frac{e^{-\mu_0 t}}{(1-\kappa_1 u_n t)^{1/2}(1-\kappa_2 u_n t)^{1/2}} dt\right)^{-1} \quad (2.18)$$

and

$$\int_0^{\hat{t}} \frac{e^{-\mu_0 t}}{(1-\kappa_1 u_n t)^{1/2}(1-\kappa_2 u_n t)^{1/2}} dt = \frac{1}{\beta_0 a(0)}. \quad (2.19)$$

Recall that t^* in (2.18) is the smallest positive root of (2.8). For a converging wave for which $|a(0)| > \gamma$, the time $\hat{t}$ is the time required for a shock wave to form. The time t^* is the time required for a *caustic* to form, i.e., the time required for the wavefront to intersect itself. Theorem 1(i) shows that when $|a(0)| > \gamma$ the shock forms before the caustic. Theorem 1 (ii) shows that when $|a(0)| < \gamma$ the shock does not occur but the caustic does. The case $|a(0)| = \gamma$ corresponds to the simultaneous formation of a shock and a caustic. Observe, also, from (2.18) that for a converging wave we can have $\mu_0 \leq 0$ and still have a nonzero critical initial amplitude.

By (2.18), we can easily establish that if γ is not zero, then

$$\frac{\partial \gamma}{\partial \mu_0} > 0, \qquad \mu_0 \neq 0. \quad (2.20)$$

Finally, we should remark that Bailey and Chen's asymptotic results for case when $|a(0)| > \alpha$ are also applicable to converging waves. We simply need to replace α by γ and t_∞ by $\hat{t}$.

ACKNOWLEDGMENT

This work was supported by the U.S. Atomic Energy Commission.

1 P. B. Bailey and P. J. Chen, Arch. Ratl. Mech. Anal. **41**, 121 (1971).
2 P. B. Bailey and P. J. Chen, Arch. Ratl. Mech. Anal. **44**, 212 (1972).
3 E. L. Ince, *Ordinary Differential Equations* (Longmans, Green, New York, 1926).
4 See Sec. 4 of Refs. 1 and 2.
5 Reference 1, Eq. (4.13).
6 Reference 1, Eq. (4.14).
7 B. D. Coleman and M. E. Gurtin, Arch. Ratl. Mech. Anal. **19**, 239 (1965).
8 B. D. Coleman, J. M. Greenberg, and M. E. Gurtin, Arch. Ratl. Mech. Anal. **22**, 333 (1966).
9 Reference 3, Secs. 2.13 and 2.14.
10 Theorem 4.3 of Ref. 1.
11 T. Y. Thomas, J. Math. Mech. **6**, 455 (1957).
12 E. Varley and J. Dunwoody, J. Mech. Phys. Solids **13**, 17 (1965).
13 E. Varley and E. Cumberbatch, J. Inst. Math. Appl. **1**, 101 (1965).
14 E. Varley, Arch. Ratl. Mech. Anal. **19**, 215 (1965).
15 P. J. Chen, J. Acoust. Soc. Amer. **43**, 982 (1968).
16 P. J. Chen, Arch. Ratl. Mech. Anal. **30**, 81 (1968).
17 P. J. Chen, Arch. Ratl. Mech. Anal. **31**, 228 (1968).
18 M. L. Doria and R. M. Bowen, Phys. Fluids **13**, 867 (1970).
19 E. P. Lane, *Metric Differential Geometry of Curves and Surfaces* (U. Chicago Press, Chicago, 1940).
20 Reference 2.

Harmonic Analysis of Analytic Functions on Hyperspheres

Brian Lee Beers
Department of Physics and Astronomy, Southern Illinois University, Carbondale, Carbondale, Illinois 62901
(Received 17 December 1971)

The real analytic functions on the hypersphere S^n are shown to be in one-to-one correspondence with the family of series of hyperspherical harmonics with exponentially falling coefficients. These functions may be continued onto a larger complex manifold on which they represent holomorphic functions. The convergence of the harmonic expansions for the real analytic functions on S^n is governed by the singularity structure of the continued function on this complex manifold.

INTRODUCTION

In a recent paper[1] the convergence properties of series of spherical harmonics and Wigner functions were explored. The following facts emerged: The family of globally real analytic functions is in one-to-one correspondence with the family of series with exponentially falling coefficients; these same series may be continued into a region of the complexified manifold; the domain of convergence in the complexified domain is determined by the singularity structure of the function; the coefficients of the expansion may be determined by certain integrals over the complex manifold.

Series of spherical harmonics represent functions on the sphere in three dimensions S^2, while Wigner series represent functions on the sphere in four dimensions S^3 [$SU(2)$ is equivalent to S^3 as a manifold]. In addition, the above properties are well known[2] for functions on the unit circle, i.e., Fourier series. With the exception of the question of integral representations, which will be treated elsewhere, it is the purpose of the present note to generalize these results to hyperspheres of arbitrary dimension S^n.

The work in I was motivated by a desire to understand the convergence properties of partial wave expansions for production amplitudes. As an extension of this understanding, it would be desirable to know the convergence properties of expansions of these amplitudes on phase space.[3] N-particle nonrelativistic phase space is equivalent to a sphere in $3N$-3 dimensions, leading directly to the present problem.

The approach taken here differs considerably from that of I in that it relies on the embedding of the manifolds in Euclidean space and makes heavy use of potential theory. While the intrinsic approach of I is mathematically more desirable, the restatement of the problem in the more familiar language of potenas $t \to \infty$ whenever κ_1 and κ_2 are not zero. In applying (2.16) and (2.17) it is important to note that the requirement $|a(0)| < \alpha$ requires α to be nonzero. Thus, μ_0 must obey the condition (2.11).

Case II: Converging Waves

Theorem 1 shows that when $\operatorname{sgn} a(0) = \operatorname{sgn}\beta_0$ and $|a(0)| > \gamma$, then $\lim_{t\to\hat{t}} |a(t)| = \infty$, where, by (2.1) and (2.2), γ and $\hat{t}$ are given by

$$\gamma = \left(|\beta_0| \int_0^{t^*} \frac{e^{-\mu_0 t}}{(1-\kappa_1 u_n t)^{1/2}(1-\kappa_2 u_n t)^{1/2}} dt\right)^{-1} \quad (2.18)$$

and

$$\int_0^{\hat{t}} \frac{e^{-\mu_0 t}}{(1-\kappa_1 u_n t)^{1/2}(1-\kappa_2 u_n t)^{1/2}} dt = \frac{1}{\beta_0 a(0)}. \quad (2.19)$$

Recall that t^* in (2.18) is the smallest positive root of (2.8). For a converging wave for which $|a(0)| > \gamma$, the time $\hat{t}$ is the time required for a shock wave to form. The time t^* is the time required for a *caustic* to form, i.e., the time required for the wavefront to intersect itself. Theorem 1(i) shows that when $|a(0)| > \gamma$ the shock forms before the caustic. Theorem 1 (ii) shows that when $|a(0)| < \gamma$ the shock does not occur but the caustic does. The case $|a(0)| = \gamma$ corresponds to the simultaneous formation of a shock and a caustic. Observe, also, from (2.18) that for a converging wave we can have $\mu_0 \le 0$ and still have a nonzero critical initial amplitude.

By (2.18), we can easily establish that if γ is not zero, then

$$\frac{\partial \gamma}{\partial \mu_0} > 0, \qquad \mu_0 \neq 0. \quad (2.20)$$

Finally, we should remark that Bailey and Chen's asymptotic results for case when $|a(0)| > \alpha$ are also applicable to converging waves. We simply need to replace α by γ and t_∞ by $\hat{t}$.

ACKNOWLEDGMENT

This work was supported by the U.S. Atomic Energy Commission.

1 P. B. Bailey and P. J. Chen, Arch. Ratl. Mech. Anal. **41**, 121 (1971).
2 P. B. Bailey and P. J. Chen, Arch. Ratl. Mech. Anal. **44**, 212 (1972).
3 E. L. Ince, *Ordinary Differential Equations* (Longmans, Green, New York, 1926).
4 See Sec. 4 of Refs. 1 and 2.
5 Reference 1, Eq. (4.13).
6 Reference 1, Eq. (4.14).
7 B. D. Coleman and M. E. Gurtin, Arch. Ratl. Mech. Anal. **19**, 239 (1965).
8 B. D. Coleman, J. M. Greenberg, and M. E. Gurtin, Arch. Ratl. Mech. Anal. **22**, 333 (1966).
9 Reference 3, Secs. 2.13 and 2.14.
10 Theorem 4.3 of Ref. 1.
11 T. Y. Thomas, J. Math. Mech. **6**, 455 (1957).
12 E. Varley and J. Dunwoody, J. Mech. Phys. Solids **13**, 17 (1965).
13 E. Varley and E. Cumberbatch, J. Inst. Math. Appl. **1**, 101 (1965).
14 E. Varley, Arch. Ratl. Mech. Anal. **19**, 215 (1965).
15 P. J. Chen, J. Acoust. Soc. Amer. **43**, 982 (1968).
16 P. J. Chen, Arch. Ratl. Mech. Anal. **30**, 81 (1968).
17 P. J. Chen, Arch. Ratl. Mech. Anal. **31**, 228 (1968).
18 M. L. Doria and R. M. Bowen, Phys. Fluids **13**, 867 (1970).
19 E. P. Lane, *Metric Differential Geometry of Curves and Surfaces* (U. Chicago Press, Chicago, 1940).
20 Reference 2.

Harmonic Analysis of Analytic Functions on Hyperspheres

Brian Lee Beers
Department of Physics and Astronomy, Southern Illinois University, Carbondale, Carbondale, Illinois 62901
(Received 17 December 1971)

The real analytic functions on the hypersphere S^n are shown to be in one-to-one correspondence with the family of series of hyperspherical harmonics with exponentially falling coefficients. These functions may be continued onto a larger complex manifold on which they represent holomorphic functions. The convergence of the harmonic expansions for the real analytic functions on S^n is governed by the singularity structure of the continued function on this complex manifold.

INTRODUCTION

In a recent paper[1] the convergence properties of series of spherical harmonics and Wigner functions were explored. The following facts emerged: The family of globally real analytic functions is in one-to-one correspondence with the family of series with exponentially falling coefficients; these same series may be continued into a region of the complexified manifold; the domain of convergence in the complexified domain is determined by the singularity structure of the function; the coefficients of the expansion may be determined by certain integrals over the complex manifold.

Series of spherical harmonics represent functions on the sphere in three dimensions S^2, while Wigner series represent functions on the sphere in four dimensions S^3 [$SU(2)$ is equivalent to S^3 as a manifold]. In addition, the above properties are well known[2] for functions on the unit circle, i.e., Fourier series. With the exception of the question of integral representations, which will be treated elsewhere, it is the purpose of the present note to generalize these results to hyperspheres of arbitrary dimension S^n.

The work in I was motivated by a desire to understand the convergence properties of partial wave expansions for production amplitudes. As an extension of this understanding, it would be desirable to know the convergence properties of expansions of these amplitudes on phase space.[3] N-particle nonrelativistic phase space is equivalent to a sphere in $3N$-3 dimensions, leading directly to the present problem.

The approach taken here differs considerably from that of I in that it relies on the embedding of the manifolds in Euclidean space and makes heavy use of potential theory. While the intrinsic approach of I is mathematically more desirable, the restatement of the problem in the more familiar language of potential theory seems intuitively pleasing. In any case, the approach is another tool for studying problems of this category, a category which clearly has not been delineated as yet. The expectation is that the theorems may be extended to compact semisimple Lie groups and their connected homogeneous spaces.

In Sec. 1 we give a brief review of facts concerning S^n, harmonic polynomials, and hyperspherical harmonics. We extend this information into the complex domain in order to obtain certain key results. Section 2 consists of a concise set of theorems concerning expansions in hyperspherical harmonics and analytic functions on S^n. An appendix considers an integral needed in the text. Except for the reliance on potential theory to establish certain key facts, the actual proof of many of the theorems relevant to S^n is identical to that in I. We refer the reader to this source for these proofs, and for introductory material concerning real and complex manifolds.

1. FACTS ABOUT S^n

Let R^{n+1} denote real Euclidean space of $n+1$ dimensions. Thus R^{n+1} consists of the column vectors $\mathbf{X}$ with entries $X_1, \ldots, X_{n+1}$, where the X_i are real. We denote the associated row vector by $\tilde{\mathbf{X}}$. The hypersphere $S^n(1) = S^n$ is the set of points satisfying

$$\tilde{\mathbf{X}}\mathbf{X} = \sum_{i=1}^{n+1} X_i^2 = 1. \tag{1.1}$$

S^n is a real analytic manifold of dimension n with the real analytic structure inherited from R^{n+1}.[4] In particular, each of the functions X_i may be considered to be a real analytic function on S^n.

The Laplacian operator on R^{n+1} is given by

$$\nabla_{n+1}^2 = \sum_{i=1}^{n+1} \frac{\partial^2}{\partial X_i^2} \tag{1.2}$$

An harmonic polynomial $H(n+1, l, \mathbf{X})$ on R^{n+1} of degree l is a homogeneous polynomial in the variables X_i of degree l which satisfies Laplace's equation

$$\nabla_{n+1}^2 H(n+1, l, \mathbf{X}) = 0. \tag{1.3}$$

Thus, each such polynomial has the form

$$H(n+1, l, \mathbf{X}) = \sum_{\{l\}} a_{l_1 \cdots l_{n+1}} X_1^{l_1} \cdots X_{n+1}^{l_{n+1}}, \tag{1.4}$$

where

$$\sum_{i=1}^{n+1} l_i = l \tag{1.5}$$

and the sum is taken over all partitions $\{l\}$ of l.

We introduce in R^{n+1} the radial distance X given by

$$X = \left(\sum_{i=1}^{n+1} X_i^2\right)^{1/2} \tag{1.6}$$

together with a maximal atlas of coordinate patches on S^n which make S^n into a real analytic manifold. As an example of a coordinate patch, we introduce the standard hyperspherical polar coordinates[5] $\theta_1, \ldots, \theta_{n-1}, \phi$ via

$$\begin{aligned} X_1 &= X\cos\theta_1, \\ X_2 &= X\sin\theta \cos\theta_2, \\ &\vdots \\ X_{n-1} &= X \sin\theta_1 \sin\theta_2 \cdots \cos\theta_{n-1}, \\ X_n &= X \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-1} \cos\phi, \\ X_{n+1} &= X \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-1} \sin\phi, \end{aligned} \tag{1.7}$$

The domain of the patch is all those points in S^n where the Jacobian of the transformation in nonzero. The measure $d\Omega_n$ on S^n is defined by the relation

$$dV_{n+1} = X^n dX d\Omega_n, \tag{1.8}$$

where dV_{n+1} is the standard Euclidean volume in R^{n+1}. In terms of hyperspherical polar coordinates $d\Omega_n$ has the form

$$d\Omega_n = (\sin\theta_1)^{n-1}(\sin\theta_2)^{n-2} \cdots (\sin\theta_{n-1}) d\theta_1 \cdots d\theta_{n-1} d\phi. \tag{1.9}$$

The total area Ω_n of S^n is readily found to be[6]

$$\Omega_n = \int_{S^n} d\Omega_n = 2(\pi)^{(n+1)/2}/\Gamma(\tfrac{1}{2}(n+1)). \tag{1.10}$$

The family of continuous functions on S^n may be converted into an Hilbert space with the introduction of the scalar product

$$(f,g)_n = \int_{S^n} f^*(\Omega) g(\Omega) d\Omega_n. \tag{1.11}$$

The associated norm is of course given by

$$\|f\|_n = +[(f,f)_n]^{1/2} \tag{1.12}$$

In terms of X and a coordinate patch on S^n, the Laplacian operator becomes

$$\nabla_{n+1}^2 = X^{-n} \frac{\partial}{\partial X}\left(X^n \frac{\partial}{\partial X}\right) + X^{-2} L^{2n} \tag{1.13}$$

Here L^{2n} is the quadratic Laplace–Beltrami differential operator on S^n. Its explicit form depends on the choice of coordinate patch, but the form depends only on the patch and not on X.

Let $H(n+1, l, \mathbf{X})$ be an harmonic polynomial on R^{n+1}. As $H(n+1, l, \mathbf{X})$ is homogeneous of order l, it may be written in the form

$$H(n+1, l, \mathbf{X}) = X^l Y(n, l, \Omega). \tag{1.14}$$

$Y(n, l, \Omega)$ is a function on S^n called a surface harmonic or hyperspherical harmonic. $Y(n, l, \Omega)$ is simply the restriction of $H(n+1, l, \mathbf{X})$ to S^n. As it is formed from finite sums of finite products of the X_i, each of which is analytic, $Y(n, l, \Omega)$ is a globally real analytic function on S^n.

By inserting the definition (1.14) into Laplace's equation (1.3) using the form (1.13), we find the important property

$$L^{2n}\{Y(n, l, \Omega)\} = -l(l+n-1) Y(n, l, \Omega). \tag{1.15}$$

Using this fact, one may readily show that surface harmonics of different degree are orthogonal on S^n, i.e.,

$$(Y(n,l), Y(n,l')) = 0, \quad l \neq l'. \tag{1.16}$$

For fixed l, there are $N(n,l)$ linearly independent harmonic polynomials, where[7]

$$N(n,l) = (2l + n - 3)(l + n - 2)!\{l!(n-1)!\}^{-1}. \quad (1.17)$$

In each such subspace of polynomials of fixed degree we may choose an orthogonal set, order them in a linear array, and label them with a single index m. These functions may then be normalized to yield

$$(Y(n,l,m), Y(n,l',m')) = \delta_{ll'}\delta_{mm'}. \quad (1.18)$$

We have the following important lemma.[8]

Lemma 1: Let Y_l be any surface harmonic of degree l. Then

$$|Y_l(\Omega)|^2 \leq \Omega_n^{-1} N(n,l) \|Y_l\|^2 \quad (1.19)$$

for all $\Omega \in S^n$.

Let us now extend these harmonic polynomials into the complex domain. Let C^{n+1} denote complex Euclidean space of $n+1$ (complex) dimensions. C^{n+1} consists of column vectors $\mathbf{Z}$ with entries $Z_1 \cdots Z_{n+1}$, where $Z_i = X_i + iY_i, X_i, Y_i$ real. C^{n+1} may also be considered as a real vector space of dimension $2n+2$. Consider a polynomial of the form (1.4) extended to C^{n+1}

$$H(n+1,l,\mathrm{Z}) = \sum_{\{l\}} a_{l_1 \cdots l_{n+1}} Z_1^{l_1} \cdots Z_{n+1}^{l_{n+1}}. \quad (1.20)$$

$H(n+1,l,\mathbf{Z})$ is an entire *complex analytic* (holomorphic) function of $\mathbf{Z}$. In addition, for $\mathbf{Y}$ fixed ($\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$), $H(n+1,l,\mathrm{Z})$ is an harmonic function on R^{n+1}. Using the Cauchy–Riemann conditions, it follows that $H(n+1,l,\mathrm{Z})$ is also an harmonic function in the $(2n+2)$-dimensional real space spanned by $\mathbf{X}$ and $\mathbf{Y}$. Each harmonic polynomial in R^{n+1} gives rise to an holomorphic harmonic polynomial in C^{n+1}. We consider the expression of these polynomials in terms of hyperspherical type coordinates in $(2n+2)$ dimensions. We define the radial distance as

$$Z^2 = \mathbf{Z}^\dagger \mathbf{Z} = \sum_{i=1}^{n+1} (X_i^2 + Y_i^2), \quad (1.21)$$

to introduce a coordinate patch on S^{2n+1}, and insert into (1.20) to yield

$$H(n+1,l,\mathbf{Z}) = Z^l Y(n,l,\Omega_Z). \quad (1.22)$$

We may now use Lemma 1 to establish the growth of the harmonic polynomials when extended to C^{n+1}. We first introduce semitoroidal coordinates on S^{2n+1} by the rule

$$Z_i = X_i e^{i\phi_i}, \quad (1.23)$$

where the X_i are given by (1.7) with the exception that X is replaced by Z. In this coordinate patch, we can readily see that

$$d\Omega_{2n+1} = \Theta\, d\Omega_n d\phi_1 \cdots d\phi_{n+1}, \quad (1.24)$$

where Θ is the product of the X_i for $Z = 1$. If we now compute the norm of $Y(n,l,\Omega_Z)$ on S^{2n+1}, we find that the integrations over ϕ_i kill Z the terms in Y^*Y except those which appear in the S^n norm. We readily establish therefore that

$$\|Y(n,l,\Omega_Z)\|_{2n+1}^2 \leq (2\pi)^{n+1} N(n,l) C \|Y(n,l,\Omega_X)\|_n^2, \quad (1.25)$$

where C is the maximum value of $|\Theta|$. Applying (1.25) and Lemma 1 with appropriate factors yields the following theorem.

Theorem 1: Let $H(n+1,l,\mathbf{X})$ be an harmonic polynomial as given in (1.4), and let $Y(n,l,\Omega_X)$ as given in (1.14) be normalized to unity. Then the analytic extension $H(n+1,l,\mathrm{Z})$ of $H(n+1,l,\mathbf{X})$ satisfies the bound

$$|H(n+1,l,\mathrm{Z})| \leq M(n,l) Z^l, \quad (1.26)$$

where

$$M(n,l) = [\Omega_{2n+1}^{-1}(2\pi)^{n+1} C\, N(2n+1,l) N(n,l)]^{1/2}. \quad (1.27)$$

Having considered the complexification of R^{n+1}, we now consider the complexification of S^n, which we call S^{nc}. S^{nc} is the set of points in C^{n+1} satisfying

$$\tilde{\mathbf{Z}}\mathbf{Z} = 1. \quad (1.28)$$

Introducing the radial distance Y in $\mathbf{Y}$ space ($\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$), this condition is equivalent to the conditions

$$X^2 - Y^2 = 1, \quad \tilde{\mathbf{X}}\mathbf{Y} = 0. \quad (1.29)$$

For fixed Y, the points in S^{nc} consist of the points in $\mathbf{X}$ and $\mathbf{Y}$ lying on spheres such that $\mathbf{X}$ and $\mathbf{Y}$ are orthogonal vectors. Varying Y gives all of S^{nc}. The spherical portion in $\mathbf{X}$ is growing from the unit sphere as the $\mathbf{Y}$ sphere grows from the origin.

S^{nc} is a complex analytic manifold of complex dimension n, with complex analytic structure inherited from C^{n+1}. This is readily seen, as S^{nc} is defined by a holomorphic function on C^{n+1}, and the composition of holomorphic functions is again holomorphic. S^{nc} is viewed as the complexification of S^n. As each of the functions Z_i is holomorphic when restricted to S^{nc}, each of the functions $H(n+1,l,\mathrm{Z})$ is an entire holomorphic function on S^{nc}.

It will be convenient to relate the distance extended into S^{2c} (away from the real portion) to the radial distance in C^{n+1}. To do this, we parameterize X and Y satisfying (1.29) by

$$X = \cosh\alpha, \quad Y = \sinh\alpha, \quad 0 \leq \alpha \leq \infty. \quad (1.30)$$

This yields

$$Z(\alpha) = (\cosh^2\alpha + \sinh^2\alpha)^{1/2} = \cosh^{1/2} 2\alpha. \quad (1.31)$$

2. HARMONIC SERIES AND ANALYTIC FUNCTIONS ON S^n AND S^{nc}

Let $B^n(X)$ denote an open ball of radius X in $n+1$ real dimensions, i.e., the set whose boundary is $S^n(X)$. We shall show in this section that there exists a one-to-one correspondence between the real analytic functions on S^n and the family of absolutely and uniformly convergent series of harmonic polynomials in $B^n(X)$ for some $X > 1$. Such series in fact may be complexified to holomorphic functions in C^{n+1}, the given series converging inside the ball $B^{2n+1}(Z = X)$. Restricting these series to S^{nc} gives a convergent expansion on those portions of S^{nc} interior to $B^{2n+1}(Z = X)$.

It is well known that the hyperspherical harmonics form an orthonormal basis for the Hilbert Space of continuous functions on S^n.[9] Consider the series

$$I_f = \sum_{l\ 0}^{\infty} \sum_{m=0}^{N(n,l)} a_l^m Y(n,l,m), \quad (2.1)$$

where

$$a_l^m = (Y(n,l,m),f). \tag{2.2}$$

Via relation (1.15), the hermiticity of L^{2n}, and the Schwarz inequality, we may readily establish the following theorem:

Theorem 2: Suppose f is $2p$ times continuously differentiable ($p = 0, 1, \cdots$). Then a_l^m in (2.2) satisfies the bound

$$|a_l^m| \le [l(l+n-1)]^{-p} \, \|L^{2n}f\|_n \tag{2.3}$$

(see Theorem 6 of I). Note that this says that infinitely differentiable functions give rise to coefficients which fall off faster than any inverse power of l. We will be able to improve this if f is analytic.

Via (2.1), we construct the series

$$I_f(\mathbf{X}) = \sum_{l=0}^{\infty} \sum_{m=0}^{N(n,l)} a_l^m X^l Y(n,l,m). \tag{2.4}$$

By using the estimates (1.19), (1.26), and (2.3), it is straight forward to prove the following theorem.

Theorem 3: Let f be a continuous function on S^n. For all $X_0 < 1$, the series (2.4) converges absolutely and uniformly to a real analytic function of $\mathbf{X}$ in the closed ball $\overline{B^n}(X_0)$. $I_f(\mathbf{X})$ is harmonic in this ball ($\nabla_{n+1}^2 I_f = 0$). Furthermore, $I_f(\mathbf{X})$ may be continued to a holomorphic function $I_f(\mathbf{Z})$ represented by the series

$$I_f(\mathbf{Z}) = \sum_{l=0}^{\infty} \sum_{m=0}^{N(n,l)} a_l^m H(n,l,m,\mathbf{Z}), \tag{2.5}$$

which converges everywhere in the closed ball $\overline{B}^{2n+1}$ ($Z = X_0$).

If we restrict $I_f(\mathbf{X})$ to a fixed value of $X_0 < 1$, $I_f(X_0)$ represents a real analytic function on the sphere of radius X_0. This follows, since I_f is real analytic in the X_i, and the transformation to X and a coordinate patch on $S^n(X)$ is necessarily analytic. The resulting function is clearly analytic in each coordinate patch. From this, it should be clear that if we begin with a continuous function on a sphere of radius $X_0 > 1$, we can generate a real analytic function on S^n. That is, we simply consider the series

$$J_f(\mathbf{X}) = \sum_{l=0}^{\infty} \sum_{m=0}^{N(n,l)} b_l^m \left(\frac{X}{X_0}\right)^l Y(n,l,m), \tag{2.6}$$

where the b_l^m are obtained as in (2.2), only the integration is over $S^n(X_0)$.

$J_f(\mathbf{X})$ represents the regular solution of Dirichlet's interior problem for the Laplace equation in R^{n+1} with continuous boundary values given on the sphere $S^n(X_0)$.[10] Such solutions give rise to real analytic functions on S^n. Do all real analytic functions on S^n arise in this manner?[11] This question is answered in the affirmative in the following theorem.

Theorem 4: Let f be a real analytic function on S^n. Let g be the regular solution of Laplace's equation $\nabla_{n+1}^2 g = 0$ in the interior of S^n which takes on the values f when approached from within. Then g is real analytic in some ball $B(X_0)$ for some $X_0 > 1$.

To prove this, we will show that the Poisson representation for the solution of the interior problem may be continued into a sphere of radius $X_0 > 1$. The solution g of the interior problem is given by[12]

$$g(X,\Omega) = \frac{(1-X^2)}{\Omega_n} \int d\Omega_n' \frac{f(\Omega')}{(1+X^2-2X\cos\gamma)^{(n+1)/2}}, \tag{2.7}$$

where γ is the angle between the observation point vector (X,Ω) and the integration point vector $(1,\Omega')$. For X,Ω such that the denominator in (2.7) does not vanish, $g(X,\Omega)$ represents an analytic function of (X,Ω). The integral may be written

$$g(X,\Omega) = \frac{1}{\Omega_n} \frac{(1-X^2)}{(2X)^{(n+1)/2}} \times \int d\Omega' \frac{f(\Omega')}{(\cos u - \cos\gamma)^{(n+1)/2}}, \tag{2.8}$$

where

$$y \equiv \cos u = \tfrac{1}{2}(X + X^{-1}). \tag{2.9}$$

The map (2.9) from the complex X plane to the complex $\cos u$ plane maps circles in X onto ellipses in y. The unit circle in X is mapped onto the interval $[-1,1]$, while the circles of radius R, $1/R$ are mapped onto the same ellipse in y, having foci at ± 1 and semimajor axis $\frac{1}{2}(R + 1/R)$

For $|X| < 1$, (2.8) represents an analytic function of X when Ω is real. We already know that $g(X,\Omega)$ is analytic interior to S^n. We also know that $g(X=1,\Omega) = f(\Omega)$ is analytic in Ω. By Hartog's theorem[13] we need only establish that g is an analytic function of X for fixed Ω in order to establish the analyticity of g.

This is accomplished as follows. Choose the observation point Ω. The axes of integration may now be chosen so that $\cos\gamma = \cos\theta_1$, where θ_1 is the first hyperspherical angle defined in Eq. (1.7). We may express $d\Omega_n$ in terms of the $(n-1)$ spherical area by the relation

$$d\Omega_n = (\sin\theta_1)^{n-1} d\Omega_{n-1}. \tag{2.10}$$

Inserting this information into Eq. (2.8) yields

$$G(X) \equiv g(X, \Omega = \Omega_0) = \frac{(1-X^2)}{(2X)^{(n+1)/2}} \times \int_{-1}^{1} \frac{(1-W^2)^{(n-2)/2}}{(y-W)^{(n+1)/2}} f'(W)\,dW, \tag{2.11}$$

where

$$f'(W) = \Omega_N^{-1} \int d\Omega_{n-1} f(W, \Omega_{n-1}) \tag{2.12}$$

and

$$W = \cos\theta_1. \tag{2.13}$$

For X in a neighborhood of 1, y is in a neighborhood of 1. The only portion of the integral (2.11) which can be singular is in a neighborhood of the endpoint $W = 1$.[14] Now $f'(W)$ has a convergent power series expansion in powers of $(1-W^2)$ near $W = 1$.

This may be seen as follows: A coordinate patch in a neighborhood of $W = 1$ consists of the variables X_i, $i > 1$. Since f is analytic it has a power series expansion in the X_i. Inserting such an expansion into Eq. (2.12) together with the definitions (1.7) yields an expansion of $f'(W)$ in powers of $(1-W^2)^{l/2}$. As $\cos\theta_i$ is odd in $[0,\pi]$ and $\sin\theta_i$ is even, only terms with l even survive the integration.

Inserting a power series expansion in $(1-W^2)$ into the integral (2.11) and explicitly evaluating the integral from $1-\delta$ to 1, it can be seen that $G(X)$ is in fact analytic in X in a neighborhood of $X = 1$ (see Appendix).

These arguments show that in a full neighborhood of any point on S^n, $g(X,\Omega)$ is a real analytic function. By using the continuity of the radius of convergence of real analytic functions and the compactness of S^n, it follows that $g(X,\Omega)$ is real analytic in $B(X_0)$ for some $X_0 > 1$. (See Theorems 7–10 of I). This proves the theorem.

Since $g(X,\Omega)$ is real analytic in the interior of $B(X_0)$, $\nabla^2_{n+1} g(X,\Omega)$ is also real analytic in the interior of $B(x_0)$. But $\nabla^2_{n+1} g(X,\Omega)$ vanishes inside S^n. By continuation, it also vanishes inside $B(X_0)$. $g(X,\Omega)$ may therefore be represented by a series of the form (2.6) which converges to $g(X,\Omega)$ everywhere in $B(X_0)$. Since the $b_l'^m$ in (2.6) are bounded by a constant, it follows that the restriction of this series to $S^n(X=1)$ is a surface harmonic expansion which has coefficients falling off exponentially:

$$|a_l^m| = |b_l^m X_0^{-l}| \le Ce^{-l \ln X_0}. \tag{2.14}$$

We now have sufficient information to state the theorems which characterize the analytic functions on S^n and their harmonic expansions. We make the following definitions in order to proceed.

Definition 1: Let I_f be a series of the form (2.1). The *modulus of convergence* Z_0 is defined by the formula

$$Z_0^{-1} = \lim \sup |a_l^m|^{1/l}, \tag{2.15}$$

where the limit superior is taken over all l, m indicated in (2.1).

Definition 2: The *open superball* of modulus Z_1, $SB^n(Z_1)$ is the set of all points in S^{nc} for which $Z(\alpha)$ is less than Z_1, where $Z(\alpha)$ is defined in (1.31).

Definition 3: The *characteristic growth function* $\Delta_l(\alpha)$ is defined as

$$\Delta_l(\alpha) = M(n,l)[Z(\alpha)]^l. \tag{2.16}$$

Note that $Z(\alpha)$ is a monotone function of α and that

$$\lim_{l\to\infty} [M(n,l)]^{1/l} \to 1. \tag{2.17}$$

Theorem 5: Let I_f be a series of the form (2.1), with $Z_0 \neq 0$ given by (2.15). Then Z_0 is the least upper bound of those Z's for which the sequence $|a_l^m|\,\Delta_l(\alpha)$ is bounded (Theorem 2 of I).

Theorem 6: I_f converges absolutely and uniformly to a holomorphic function everywhere in $SB^n(Z_0)$. $SB^n(Z_0)$ may not be extended by analytic completion as it represents an envelope of holomorphy [Theorem 3 of I]. The second portion of the theorem may be seen most easily by observing that $B^{2n+1}(Z_0)$ is an envelope of holomorphy.[15]

Theorem 7: Let f be a real analytic function on S^n. Then f has a unique extension to a holomorphic function in the interior of some maximal superball of holomorphy $SB^n(\alpha_1)$: That is, f is holomorphic in $SB^n(\alpha_1)$, and singular in any larger superball. (See Theorems 7–12 of I).

Lemma 2: Let f be real analytic on S^n, $SB(\alpha_1)$ its maximal superball of holomorphy. Let f' be the unique regular (at the origin) solution of Laplace's equation which agrees with f on S^n. Let $B^{2n+1}(Z_0)$ be the maximal ball in C^{n+1} in which f' is holomorphic. Then $Z(\alpha_1) = Z_0$.

Proof: Clearly f' agrees with f everywhere on S^{nc} inside $B^{2n+1}(Z_0)$, so the only possibility is that $Z(\alpha_1) \geqslant Z_0$. If $f(\mathbf{Z})$ is a solution of the Laplace's equation $\nabla^2_{n+1}(x) f(\mathbf{Z}) = 0$, then $f(\mathbf{Z}e^{i\varphi})$ is also a solution. But the set of points $\mathbf{Z}e^{i\varphi}$ where $\mathbf{Z}$ is restricted to S^{nc} and a fixed value of α sweeps out all points on the sphere $S^{2n+1}(Z(\alpha))$. By hypothesis we are able to continue f into a larger superball. There must exist a full neighborhood of this superball in which f' is analytic, and hence satisfies Laplace's equation. From this set of points, we can generate all the points in a ball larger than $B^{2n+1}(Z_0)$ by the transformation $\mathbf{Z} \to \mathbf{Z}e^{i\varphi}$. Since Laplace's equation must be satisfied at these points, the function is required to be analytic there, leading to a contradiction, unless $Z(\alpha_1) = Z_0$.

Via this lemma and (2.14), we readily establish the following theorem.

Theorem 8: Let f be a real analytic function on S^n, and let $SB^n(\alpha_0)$ be the maximal superball of holomorphy of its unique extension. Then, for any $Z_1 < Z(\alpha_0)$ the coefficients defined in (2.2) satisfy the inequality

$$|a_l^m| \le M(Z_1) Z_1^{-l}, \tag{2.18}$$

where $M(Z_1)$ is a constant for fixed Z_1.

Putting together our previous information we get a final complete characterization of the real analytic functions on S^n.

Theorem 9: Let f be a real analytic function on S^n and $SB^n(\alpha_0)$ the maximal superball of holomorphy of its unique analytic extension. Then the harmonic expansion for f given by Eqs. (2.1) and (2.2) converges and is holomorphic in $SB^n(\alpha_0)$. The harmonic expansion converges to the function f. Conversely, if the harmonic expansion for f has a superball of convergence $SB^n(\alpha_0)$, then f can be continued to a holomorphic function in $SB^n(\alpha_0)$ and its continuation agrees with the harmonic expansion in $SB^n(\alpha_0)$. (See Theorem 14 of I).

APPENDIX

We wish to show that the integral (2.11) is analytic in X near $X = 1$, i.e., $y \approx 1$. The only portion of the integral which can be singular is the portion near the endpoint, i.e., from 1 to $1 - \delta$. Here δ is chosen small enough so that a power series for $f'(W)$ is valid, and $1 + W$ does not vary appreciably over the interval $[1, 1-\delta]$. By expanding $f'(W)$ in powers of $(1 - W^2)$,

$$f'(W) = \sum_{l=0}^{\infty} a_l (1 - W^2)^l \tag{A1}$$

inserting into (2.11), pulling out powers of $(1 + W)$, we find that the possible singular portion of (2.11) is given by

$$J_n(X) = \frac{(1 - X^2)}{(2X)^{(n+1)/2}} \sum_{l=0}^{\infty} a_l (2)^{(n-2+2l)/2} J_n^l(X), \tag{A2}$$

where

$$J_n^l(X) = \int_{1-\delta}^{1} dW \frac{(1-W)^{(n-2+2l)/2}}{(y-W)^{(n+1)/2}}. \tag{A3}$$

By defining a quantity ϵ by

$$y = 1 + \epsilon, \tag{A4}$$

we find, using (2.9),

$$\epsilon = (X-1)^2/2X; \tag{A5}$$

changing variables in (A3) to $\mu = 1 - W$, we find

$$J_n^l(X) = \int_0^{\delta} d\mu \frac{\mu^{(n-2+2l)/2}}{(\mu+\epsilon)^{(n+1)/2}}. \tag{A6}$$

To establish the analyticity of $J_n(X)$, we must establish the analyticity of $(1-X)J_n^l(X)$. This is readily accomplished, the considerations being different as n is even or odd.

Case 1: n even. Let $n = 2k, k = 1, 2, \cdots$. By inserting this and integrating by parts, we find

$$J_{2k}^l(X) = -\left[\frac{1}{(k-\frac{1}{2})} \frac{\mu^{k+l-1}}{(\mu+\epsilon)^{k-1/2}} + \frac{(k+l-1)}{(k-\frac{1}{2})(k-\frac{3}{2})} \times \frac{\mu^{k+l-2}}{(\mu+\epsilon)^{k-(3/2)}} + \cdots + \frac{(k+l-1)!\,(\mu+\epsilon)^{l-1/2}}{(k-\frac{1}{2})(k-\frac{3}{2})\cdots(\frac{1}{2}-l)}\right]_0^{\delta}. \tag{A7}$$

The portions evaluated at $\mu = \delta$ are analytic in ϵ, and hence X, while only the last term survives at $\mu = 0$, giving $\epsilon^{l-(1/2)}$. Referring to (A5) we see that this term is also analytic near $X = 1$, except for the case $l = 0$ in which case there is a simple pole. To find the behavior of $J_n(X)$, however, we must multiply $J_n^l(X)$ by $(1-X)$, killing the pole at $X = 1$, and keeping $J_n(X)$ analytic at $X = 1$.

Case 2: n odd. Let $n = 2k+1, k = 1, 2, \cdots$. By inserting this, and integrating by parts, we find

$$J_{2k+1}^l(X) = -\left[\frac{1}{k} \frac{\mu^{k+l-1/2}}{(\mu+\epsilon)^k} + \frac{(k+l-\frac{1}{2})\mu^{k+l-3/2}}{k(k-1)(\mu+\epsilon)^{k-1}} + \cdots + \frac{(k+l-\frac{1}{2})(k+l-\frac{3}{2})\cdots(l+\frac{3}{2})}{k!} \times \frac{\mu^{l+(1/2)}}{(\mu+\epsilon)}\right]_0^{\delta} - \frac{(k+l-\frac{1}{2})(k+l-\frac{3}{2})\cdots(l+\frac{3}{2})(l+\frac{1}{2})}{k!} \times K_l(X), \tag{A8}$$

where

$$K_l(X) = \int_0^{\delta} \frac{\mu^{l-1/2}}{\mu+\epsilon} d\mu. \tag{A9}$$

By inspection, the only portion of $J_{2k+1}^l(X)$ which can be nonanalytic near $\epsilon \approx 0$ is the term with $K_l(X)$, whose behavior we now discuss.

Changing variables to $v = \mu^{1/2}$, we have

$$K_l(X) = 2\int_0^{\delta^{1/2}} \frac{v^{2l}}{v^2+\epsilon} dv. \tag{A10}$$

By using the functions $\log[(Z-\delta^{1/2})/Z]$, analytic in the plane cut from 0 to $\delta^{1/2}$, and the discontinuity relation on the cut

$$\log \frac{Z-\delta^{1/2}}{Z}\Bigg|_{Z=v-i0}^{Z=v+i0} = 2\pi i, \tag{A11}$$

(A10) may be converted into a contour integral around the cut $[0, \delta^{1/2}]$ and avoiding the poles. Thus

$$K_l(X) = \frac{i}{\pi} \oint \frac{Z^{2l}}{(Z+i\epsilon^{1/2})(Z-i\epsilon^{1/2})} \log \frac{Z-\delta^{1/2}}{Z} dZ. \tag{A12}$$

Here, the poles at $\pm i\epsilon^{1/2}$ are outside the contour. We now pass to a larger contour encircling the poles and cut by picking up the poles explicitly. The integral over the larger contour is analytic in ϵ, since as $\epsilon \to 0$ the contour is never encountered. Thus, only the pole portions may be singular in ϵ. Doing this explicitly, we have

$$[K_l(X)]_{\text{sing}} = (-1)^{l+1}\epsilon^{l-(1/2)}\left[i\log\left(\frac{i\epsilon^{1/2}-\delta^{1/2}}{i\epsilon^{1/2}+\delta^{1/2}}\right)\right]. \tag{A13}$$

Again, the only nonanalytic term is when $l = 0$, and this pole is again killed when multiplied by $(1-X)$. The logarithmic portion in brackets is real and analytic near $\epsilon \approx 0$.

ACKNOWLEDGMENTS

The central ideas of this note were formulated while the author was at the University of Maryland. I would like to thank Dr. A. J. Dragt for his invaluable assistance and encouragement during this time. This work was supported in part by the U.S. Air Force Office of Scientific Research under Grant AFOSR No. 68-1453A.

---

1 B. L. Beers and A. J. Dragt, J. Math. Phys. **11**, 2313 (1970), hereafter referred to as I.
2 E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, England, 1963), p. 161.
3 A classification of three-body states with proper behavior under rotations is given by A. J. Dragt, in J. Math. Phys. **6**, 533, 1621 (1965). A discussion of the phase space expansion for this case may be found in B. L. Beers, thesis (University of Maryland, 1970) (unpublished). Note that the results of the present note are independent of any special classification scheme.
4 L. Auslander and R. E. MacKenzie, *Introduction to Differentiable Manifolds* (McGraw-Hill, New York, 1963), p. 36.
5 *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. II, p. 233.
6 Reference 5, p. 234.
7 Reference 5, p. 237.
8 C. Muller, *Spherical Harmonics*, Lecture Notes in Mathematics, Vol. 17 (Springer-Verlag, Berlin, 1966), p. 14.
9 Reference 8, p. 40.
10 This may be verified by expansion of the Poisson kernel in a harmonic series.
11 It is true that there is some analytic function in a neighborhood of S^n which agrees with any given analytic function on S^n. In fact, there is an equivalence class $\{f\}$ with an infinite number of functions which agree with f on S^n. We have chosen a particularly simple set of functions with our choice of harmonic series. The question is, "Does each equivalence class $\{f\}$ contain a harmonic series?."
12 L. L. Helms, *Introduction to Potential Theory* (Wiley, New York, 1969), p. 16.
13 B. A. Fuks, *Theory of Analytic Functions of Several Complex Variables* (American Mathematical Society, Providence, R.I., 1963), p. 21.
14 End point analysis of this type is central to the study of analyticity of Feynman amplitudes. A summary of the techniques may be found in, R. J. Eden, P. V. Landshoff, D. T. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).
15 Reference 13, p. 195.

Structure of the Gravitational Field at Spatial Infinity

Robert Geroch
The Enrico Fermi Institute, University of Chicago, Chicago, Illinois, 60637
(Received 4 January 1972)

A formalism is introduced for analyzing the structure of the gravitational field in the asymptotic limit at spatial infinity. Consider a three-dimensional surface S in the space–time such that the initial data on S is asymptotically flat in an appropriate sense. Using a conformal completion of S by a single point Λ "at spatial infinity," the asymptotic behavior of fields on S can be described in terms of local behavior at Λ. In particular, the asymptotic behavior of the initial data on S defines four scalars which depend on directions at Λ. Since there is no natural choice of a surface S in a space–time, the dependence of these scalars on S is essential. The asymptotic symmetry group at spatial infinity, whose elements represent transformations from S to other asymptotically flat surfaces, is introduced. It is found that this group, which emerges initially as an infinite-dimensional generalization of the Poincaré group, can be reduced to the Lorentz group. A set of evolution equations is obtained: These equations describe the behavior of the four scalars under the action of the asymptotic symmetry group. The four scalars can thus be considered as fields on a three-dimensional manifold consisting of all points at spatial infinity. The notion of a conserved quantity at spatial infinity is defined, and, as an example, the expression for the energy-momentum at spatial infinity is obtained.

1. INTRODUCTION

There is available in special relativity a richer and more natural description of closed systems than in general relativity. This difference arises, essentially, because of the existence of the Poincaré group, as the group of symmetries on Minkowski space. Consider a system initially characterized locally by certain tensor fields. The action of the Poincaré group in special relativity permits a comparison of tensors at different points of the space–time manifold; hence, one can add together local contributions from various parts of the system to obtain quantities characteristic of the system as a whole. Of course, constructions of this type are unavailable in the presence of curvature. Thus, for example, it is meaningful to speak of the total energy, momentum, and angular momentum of a system in special relativity, while, *a priori*, such quantities are not well defined in general relativity.

Is there any hope at all, then, of obtaining even an incomplete global description of a gravitating system? Suppose that the metric of space–time is asymptotically flat, that is, that the metric g_{ab} approaches a Minkowski metric, in an appropriate sense, at great distances from the sources. Then the curvature would become less and less important as one moves away from the sources toward infinity. It would seem reasonable to expect in this case that, by a suitable limiting procedure, one could indeed define certain global quantities in terms of the asymptotic behavior of various fields (e.g., electromagnetic, gravitational, etc.). Of course, the asymptotic region would be well separated from the source region, and so it may be impossible to express such asymptotic quantities directly as integrals over the sources.

There are two distinct regimes in which the asymptotic behavior of the gravitational field has been found to yield useful information concerning the structure of a gravitating system.

The first of these is at null infinity, i.e., in the limit as one moves away from the sources along null geodesics Asymptotic structure in this regime has been discussed by Bondi, Van den Berg, and Metzner[1] and by Sachs.[2] A set of conditions for asymptotic flatness at null infinity have been formulated and an expression obtained for the total energy (of sources plus gravitational field) in terms of the asymptotic behavior of the gravitational field. Since the characteristic surfaces of gravitational radiation are null surfaces,[3] one would expect to see, at null infinity, the radiation emitted by the sources. This turns out to be the case. In fact, the total energy, measured at null inifinity, decreases with time at a rate depending on the flux of radiation escaping between successive null surfaces. This treatment of asymptotic structure was later reformulated by Penrose,[4] who, by means of a conformal transformation on the metric, introduced a three-dimentional surface $\mathscr{I}$ "at null infinity." Asymptotic properties of a system can be expressed directly, without the use of limiting procedures, as local structure on $\mathscr{I}$.

Alternatively, one can study asymptotic structure at spatial infinity. This topic has been discussed in a series of papers by Arnowitt, Deser, and Misner.[5] As in the null case, a set of conditions for asymptotic flatness has been formulated and an expression obtained for the total energy-momentum[6] in terms of the asymptotic behavior of the gravitational field. (In the special case when the space–time is static, one can do much more, e.g., define multipole moments.[7])

The purpose of this paper is to introduce a "local" description of the asymptotic structure of the gravitational field at spatial infinity. Our treatment represents a reformulation (more in the spirit of Penrose) of the work of Arnowitt, Deser, and Misner, as well as an extension of that work. Let M be a four-dimensional manifold with a smooth metric g_{ab} of signature $(-, +, +, +)$, for which the sources vanish outside some world tube.[8] Let S be a smooth, three-dimensional, spacelike surface in M. The induced metric on S and the extrinsic curvature of S represent the initial data for the gravitational field.[9] Asymptotic flatness of such an initial data set is interpreted to mean the existence of a conformal completion of S by a single point Λ "at spatial infinity" which displays certain features of the standard conformal completion of Euclidean 3-space. In addition, one imposes conditions on the behavior of the extrinsic curvature near Λ. Asymptotic structure of the gravitational field is then to be expressed as structure at the point Λ, asymptotic quantities such as energy and momentum as tensors at Λ, etc.

In Sec. 2, in order to illustrate the technique, we outline the asymptotic description of the electromagnetic field in flat space.

In Sec. 3 we define the notion of asymptotic flatness of an initial data set. A certain collection of tensors, defined at the point Λ, characterize the asymptotic gravitational field. An important feature of these tensors is their dependence on the surface S. If two nearby surfaces S and S' are both asymptotically flat, then the asymptotic behavior of S' relative to S can be expressed in terms of two additional tensors at Λ.

(For example, if S and S' were 3-planes in Minkowski space, then the two tensors would represent, respectively, the infinitesimal boost and the infinitesimal time translation which, applied in succession, carry S to S'.) We derive the evolution equations, which express the asymptotic field with respect to S' in terms of that with respect to S and the two tensors which fix the relationship between S and S'. Thus, because of the mere existence of such evolution equations, the asymptotic field at spatial infinity is deterministic. Physically, this determinism reflects the fact that radiation cannot escape between successive asymptotically flat 3-surfaces in the space-time.[3] (That is, such surfaces are Cauchy surfaces.[10]) More generally, nothing which the system does at any finite time can make its influence felt at spatial infinity. This situation contrasts sharply with that at null infinity.

Let S and S' be nearby asymptotically flat 3-surfaces in the space-time M. It turns out, unfortunately, that the two tensors at Λ which fix the asymptotic relationship of S' to S depend also on directions at Λ. Consequently, the group generated by such pairs of tensors (the group in curved space which replaces the Poincaré group in flat space) is infinite-dimensional. This group—the asymptotic symmetry group at spatial infinity—is discussed in Sec. 4. It turns out to be rather similar in structure to the Poincaré group, but with the translation subgroup replaced by a much larger subgroup. (The same phenomenon occurs at null infinity.[2]) It is not difficult to see intuitively why this enlargement of the translation subgroup is necessary. The general infinitesimal Poincarè transformation (i.e., the general Killing vector field) in Minkowski space has the form $\xi^a = F^a{}_b x^b + \underset{0}{\xi}{}^a$, where F_{ab} is a constant skew tensor field, $\underset{0}{\xi}{}^a$ is a constant vector field, and x^b is the position vector relative to some origin. The term $F^a{}_b x^b$, which grows without bound at infinity, represents a boost and rotation, while the constant term $\underset{0}{\xi}{}^a$ represents a translation. Suppose now that the space-time is only asymptotically flat. We wish to introduce vector fields ξ^a, which behave as Killing fields in the limit at infinity. Although no exact decomposition of ξ^a as above will be possible in the presence of curvature, we may still require that ξ^a have the same behavior as a Killing vector in Minkowski space to highest order in position. That is, we may require that ξ^a be "asymptotically linear in its position dependence." The curvature will, however, cause ambiguity in separating a unique "constant part" of ξ^a from the much larger "linear part." This ambiguity will not disappear even in the asymptotic limit, for, although the curvature becomes less and less at infinity, the domination of the constant part by the linear part becomes greater and greater. Consequently, one must admit as asymptotic Killing fields all vector fields which are asymptotically linear in position (with coefficients which form a skew tensor). In particular, the translation subgroup must encompass all vector fields which, when added to such a ξ^a, do not disturb its asymptotic behavior to highest order. The collection of such vector fields in curved space is, of course, much larger than the collection of constant vector fields in flat space. Hence, one is forced to enlarge the translation subgroup.

The fact that the asymptotic symmetry group in curved space does not coincide with the Poincaré group leads to a serious problem. Since the gravitational field at spatial infinity is defined in terms of a spacelike surface S, and since the asymptotic symmetry group determines the possible nearby surfaces to S, it is the full asymptotic symmetry group which acts on the field at spatial infinity. But global quantities, such as energy and momentum, in special relativity form representations of the Poincaré group, and it is this action of the Poincaré group which provides the usual interpretations of such quantities. Hence, it would be rather difficult, *a priori*, to interpret quantities defined in terms of the asymptotic gravitational field. However, it is found in Sec. 4 that certain combinations of the asymptotic field tensors transform, under the action of the asymptotic symmetry group, in a way independent of the generalized translations. One thus restores an action of the Lorentz group. By limiting consideration to these combinations, the problems associated with the asymptotic symmetry group are effectively eliminated.

In Sec. 5, we introduce a three-dimensional manifold $\mathcal{S}$ consisting of all points "at spatial infinity." The asymptotic gravitational field is represented by certain tensor fields on $\mathcal{S}$. By taking appropriate averages, we define conserved quantities at spatial infinity.[11] Since no influence of the sources at any finite time can affect the structure at spatial infinity, these conserved quantities can be interpreted as describing the system in the limit of the distant past. Certain of these, such as energy and momentum, can be considered as applicable to the system for all times, while others refer to the radiation emitted by the sources in the distant past. We merely write down the expression for the energy-momentum of Arnowitt, Deser, and Misner. Additional conserved quantities and their interpretation will be discussed in a subsequent paper.

2. AN EXAMPLE: ELECTROMAGNETISM

Our approach to the asymptotic structure of the gravitational field will be based on a number of geometrical constructions. It is convenient to discuss these constructions initially in a familiar setting, without the additional complications inherent in the gravitational case. Furthermore, we require an example of asymptotic structure in flat space in order to illustrate the modifications caused by curvature. For these reasons, we begin with a brief discussion of the asymptotic structure of the electromagnetic field at spatial infinity.

Let M, g_{ab} be Minkowski space, and let $F_{ab} (= F_{[ab]})$ be a solution of Maxwell's equations with charge-current J_a:

$$\nabla^b F_{ab} = J_a, \tag{1}$$

$$\nabla_{[a} F_{bc]} = 0. \tag{2}$$

Let S be a spacelike 3-plane in Minkowski space, ξ^a its unit normal, and h_{ab} its induced (positive-definite) metric (so $h_{ab} = g_{ab} + \xi_a \xi_b$). The electric and magnetic fields on S are defined by

$$E_a = F_{ab} \xi^b, \tag{3}$$

$$B_a = \tfrac{1}{2} \epsilon_{abcd} F^{bc} \xi^d. \tag{4}$$

Denote by D_a the derivative on S. Contracting (1) with ξ^a, and using (3), we obtain

$$D^a E_a = -\xi^a J_a. \tag{5}$$

Similarly, contracting (2) with $\epsilon^{abcd}\xi_d$ and using (4),

$$D^a B_a = 0. \tag{6}$$

Equations (5) and (6) are the constraint equations.

Let S' be a second 3-plane which differs infinitesimally from S. This S' can be described as follows. Each point of S' is obtained by moving a distance $\epsilon\varphi(x)$ along the normal to S at the point x of S, where $\varphi(x)$ is a scalar field on S and $\epsilon \ll 1$ is a constant. Thus, if $\varphi(x)$ is a constant, S' differs from S by an infinitesimal time translation. If, on the other hand, $\varphi(x)$ is linear, i.e., $\varphi(x) = s_a x^a$, where x^a is the position vector of x relative to some origin O on S and s_a is a constant vector field on S, then S' differs from S by an infinitesimal boost about O. But, since S and S' are both 3-planes in Minkowski space, the general admissible $\varphi(x)$ must represent some combination of these two Poincaré transformations. That is, $\varphi(x)$ must be of the form

$$\varphi(x) = s_a x^a + s. \tag{7}$$

We now obtain the evolution equations. Let a dot over a field on S denote ϵ^{-1} times the difference between the value of that field on S' and on S. Contracting (1) with h^a_c and (2) with ξ^a, we obtain

$$\dot{E}_a = \epsilon_{abc} D^b(\varphi B^c) - \varphi J_a - \varphi \xi_a(\xi^m J_m), \tag{8}$$

$$\dot{B}_a = -\epsilon_{abc} D^b(\varphi E^c), \tag{9}$$

respectively, where $\epsilon_{abc} = \epsilon_{abcd}\xi^d$ is the alternating tensor on S. Equations (8) and (9) express both the time evolution of the electromagnetic field and its behavior under boosts. Note that, applying a dot to both sides of (5) and (6), using (8), (9), and the conservation of J_a, we obtain an identity in each case.

We now return to the flat 3-space S. It is well known that Euclidean space can be conformally completed by a point at infinity. Set

$$\tilde{h}_{ab} = \Omega^2 h_{ab}, \tag{10}$$

where $\Omega = r^{-2}$ and r is the Euclidean distance from the origin O on S. It is possible to affix a single point Λ to S to obtain a new 3-manifold $\tilde{S} = S \cup \Lambda$, such that the metric $\tilde{h}_{ab}$ is smooth at Λ. (Topologically, $\tilde{S}$ is a 3-sphere.) We regard Ω as a scalar field on $\tilde{S}$ and note that the asymptotic behavior of $\Omega(\sim r^{-2})$ is characterized by the conditions

$$\Omega = 0, \quad \tilde{D}_a\Omega = 0, \quad \tilde{D}_a\tilde{D}_b\Omega = 2\tilde{h}_{ab} \tag{11}$$

at Λ, where $\tilde{D}_a$ is the derivative on $\tilde{S}$ with respect to $\tilde{h}_{ab}$.

Thus, the study of asymptotic properties of fields on S reduces to the study of local properties of the fields at Λ.

We next apply the conformal transformation (10) to the Maxwell equations. The indices of tensors with a tilde will be raised and lowered with $\tilde{h}_{ab}$ and its inverse $\tilde{h}^{ab}$. Assign to E_a and B_a dimensions[12] $\sec^{-1}$:

$$\tilde{E}_a = E_a, \quad \tilde{E}^a = \Omega^{-2}E^a,$$
$$\tilde{B}_a = B_a, \quad \tilde{B}^a = \Omega^{-2}B^a. \tag{12}$$

By using (10) and (12), the constraint equations (5) and (6) take the form (Appendix B)

$$\tilde{D}_a\tilde{E}^a = \Omega^{-1}\tilde{E}^a\tilde{D}_a\Omega,$$
$$\tilde{D}_a\tilde{B}^a = \Omega^{-1}\tilde{B}^a\tilde{D}_a\Omega, \tag{13}$$

where we have dropped the source term. (We shall be concerned with these equations only near Λ, where we shall assume $J_a = 0$.) It is convenient to assign to the function φ dimensions[12] sec:

$$\tilde{\varphi} = \Omega\varphi. \tag{14}$$

The evolution equations (8) and (9) become (Appendix B)

$$\dot{\tilde{E}}_a = \tilde{\epsilon}_{abc}\tilde{D}^b(\tilde{\varphi}\tilde{B}^c) - \Omega^{-1}\tilde{\varphi}\tilde{\epsilon}_{abc}\tilde{B}^c\tilde{D}^b\Omega,$$
$$\dot{\tilde{B}}_a = -\tilde{\epsilon}_{abc}\tilde{D}^b(\tilde{\varphi}\tilde{E}^c) + \Omega^{-1}\tilde{\varphi}\tilde{\epsilon}_{abc}\tilde{E}^c\tilde{D}^b\Omega, \tag{15}$$

where $\tilde{\epsilon}_{abc} (= \Omega^3\epsilon_{abc})$ is the alternating tensor with respect to $\tilde{h}_{ab}$, and where we have again dropped the source terms.

We wish to study Eqs. (13) and (15) asymptotically, i.e., we wish to take their limits at the point Λ. It turns out, however, that such limits will exist in general only if we permit the limit to depend on the direction of approach to Λ. Let $\tilde{T}^{a\cdots c}{}_{b\cdots d}$ be a smooth tensor field on $S = \tilde{S} - \Lambda$. We say that $\tilde{T}^{a\cdots c}{}_{b\cdots d}$ *defines a direction-dependent tensor* at Λ if the limit of $\tilde{T}^{a\cdots c}{}_{b\cdots d}$ along any smooth curve γ with endpoint Λ exists, and depends only on the unit tangent vector η^a to γ at Λ. We write the limit as follows:

$$\mathbf{T}^{a\cdots c}{}_{b\cdots d}(\eta) = \lim \tilde{T}^{a\cdots c}{}_{b\cdots d}. \tag{16}$$

Thus, $\mathbf{T}^{a\cdots c}{}_{b\cdots d}(\eta)$ is a function from the 2-sphere of unit vectors η^m at Λ to tensors at the point Λ. For example, it follows from (11) that

$$\eta_a = \lim \tfrac{1}{2}\Omega^{-1/2}\tilde{D}_a\Omega. \tag{17}$$

There is a derivative operator on direction-dependent tensors at Λ. If $\mathbf{T}^{a\cdots c}{}_{b\cdots d}(\eta)$ is such a tensor, we define $\partial_m\mathbf{T}^{a\cdots c}{}_{b\cdots d}$ to be the direction-dependent tensor obtained by taking the derivative of $\mathbf{T}^{a\cdots c}{}_{b\cdots d}(\eta)$ with respect to the unit vector η^m. Thus, the derivative satisfies the Liebnitz rule, commutes with contractions, and obeys

$$\eta^m\partial_m\mathbf{T}^{a\cdots c}{}_{b\cdots d} = 0, \tag{18}$$

for any $\mathbf{T}^{a\cdots c}{}_{b\cdots d}(\eta)$. For example, we have

$$\partial_a\eta_b = \mathbf{h}_{ab} - \eta_a\eta_b,$$
$$\partial_a\mathbf{h}_{bc} = 0, \tag{19}$$

where $\mathbf{h}_{ab}$ is the direction-dependent tensor at Λ defined by $\tilde{h}_{ab}$. Finally, the result of commuting derivatives is

$$\partial_{[m}\partial_{n]}\mathbf{T}^{a\cdots c}{}_{b\cdots d} = \eta_{\ [m}\partial_{n]}\mathbf{T}^{a\cdots c}{}_{b\cdots d}. \tag{20}$$

We shall be concerned with Maxwell fields which are asymptotically well behaved in the sense that

$$\mathbf{E}_a(\eta) = \lim\tilde{E}_a, \quad \mathbf{B}_a(\eta) = \lim\tilde{B}_a \tag{21}$$

define direction-dependent tensors at Λ. The two objects $\mathbf{E}_a$ and $\mathbf{B}_a$ characterize the asymptotic electromagnetic field.

Taking the limits of the constraint equations (13) at Λ, we obtain immediately

$$\partial_a \mathbf{E}^a = 2\eta_a \mathbf{E}^a, \quad \partial_a \mathbf{B}^a = 2\eta_a \mathbf{B}^a. \tag{22}$$

In order to write the asymptotic form of the evolution equations, we must describe the asymptotic behavior of φ. It follows from (7) and (14) that

$$\tilde{\varphi} = 0, \quad \tilde{D}_a \tilde{\varphi} = s_a, \quad \tilde{D}_a \tilde{D}_b \tilde{\varphi} = 2s\tilde{h}_{ab} \tag{23}$$

at Λ. That is to say, the value of $\tilde{\varphi}$ at Λ is zero, the derivative of $\tilde{\varphi}$ at Λ gives the infinitesimal boost associated with φ, and the second derivative of $\tilde{\varphi}$ at Λ is a multiple of $\tilde{h}_{ab}$, the constant of proportionality giving the infinitesimal time translation associated with φ. Thus, φ is completely and uniquely characterized by the number s and the vector s_a at Λ. To represent the vector s_a, we introduce the direction-dependent scalar

$$\boldsymbol{\varphi}(\eta) = \lim \Omega^{-1/2} \tilde{\varphi} = s_a \eta^a \tag{24}$$

at Λ. The fact that $\boldsymbol{\varphi}(\eta)$ is linear in η can be expressed by the equation[13]

$$\partial_{(a}\partial_{b)}\boldsymbol{\varphi} = -\boldsymbol{\varphi}\mathbf{h}_{ab} + \boldsymbol{\varphi}\eta_a\eta_b - \eta_{(a}\partial_{b)}\boldsymbol{\varphi}. \tag{25}$$

[See (19).] Taking the limits of (15) at Λ and using (24), we obtain

$$\begin{aligned} \dot{\mathbf{E}}_a &= \boldsymbol{\epsilon}_{abc}\partial^b(\boldsymbol{\varphi}\mathbf{B}^c) - \boldsymbol{\varphi}\boldsymbol{\epsilon}_{abc}\eta^b\mathbf{B}^c, \\ \dot{\mathbf{B}}_a &= -\boldsymbol{\epsilon}_{abc}\partial^b(\boldsymbol{\varphi}\mathbf{E}^c) + \boldsymbol{\varphi}\boldsymbol{\epsilon}_{abc}\eta^b\mathbf{E}^c, \end{aligned} \tag{26}$$

where $\boldsymbol{\epsilon}_{abc}$ is the direction-dependent tensor at Λ defined by $\tilde{\epsilon}_{abc}$. Note that the "time-translation part of φ," i.e., the number s in (23), does not appear in (26). That is to say, the asymptotic electromagnetic field is invariant under time translations.

To summarize, the asymptotic electromagnetic field is defined by a pair $\mathbf{E}_a(\eta)$, $\mathbf{B}_a(\eta)$, of direction-dependent vectors at Λ, subject to (22). To describe the evolution of these fields, one must first specify what boost is to be applied, i.e., one must first specify a direction-dependent scalar $\boldsymbol{\varphi}(\eta)$, subject to (25). The behavior of $\mathbf{E}_a$ and $\mathbf{B}_a$ under this $\boldsymbol{\varphi}(\eta)$ is given by (26).[14]

Suppose now that we apply to $\tilde{S}$ a further conformal transformation with conformal factor ω, where ω is smooth at Λ. To interpret this transformation geometrically, note that it amounts to choosing for the conformal factor in (10) $\omega\Omega$ rather than Ω. But we earlier set $\Omega = r^{-2}$, where r is the Euclidean distance from some origin O in S. Thus, a conformal transformation on S corresponds to a change in the choice of origin in S. In other words, the behavior of the asymptotic fields under space translations is represented at Λ by their behavior under conformal transformations on S.

In order that (11) be preserved under the application of the conformal factor ω, it is necessary that $\omega = 1$ at Λ. In fact, the ω which represents a change in origin from O to O' is $\omega = 1 + 2r^{-2}u_a x^a$, to first order in the position vector u_a of O' relative to O. Thus, $\omega = 1$ at Λ, while the derivative of ω at Λ specifies the space translation represented by ω. In particular, (21) and (12) imply that the asymptotic fields $\mathbf{E}_a$ and $\mathbf{B}_a$ are invariant under such conformal transformations, while (24) and (14) imply that $\boldsymbol{\varphi}(\eta)$ is invariant. In fact, the only change occurs in the number s of (23), to which there is added a multiple of $s_a u^a$. This behavior, of course, reflects the fact that the resolution of a Poincaré transformation into a boost and a time translation depends on a choice of origin. Since s does not appear in the evolution equations (26), all our equations at Λ are invariant under conformal transformations.

Thus, it is only the boosts and rotations, and not the time or space translations (i.e., only the Lorentz group and not the Poincaré group) which acts on the asymptotic electromagnetic field.

Since $\mathbf{E}_a(\eta)$ and $\mathbf{B}_a(\eta)$ are unaffected by asymptotic time translations, any quantity at Λ constructed from $\mathbf{E}_a$ and $\mathbf{B}_a$ is, in a sense, "conserved." Thus, if the term "conserved quantity" is to be a useful one, we must redefine it so as to require more than merely invariance under asymptotic time translations. Note that $\mathbf{E}_a$ and $\mathbf{B}_a$ define an infinite-dimensional representation of the Lorentz group. This representation is reducible. By a *conserved quantity* we shall understand any (irreducible) finite-dimensional representation of the Lorentz group obtained from the $\mathbf{E}_a$, $\mathbf{B}_a$ representation. This definition is motivated by the fact that conserved quantities in special relativity are always finite-dimensional representations.

Asymptotic conserved quantities can be obtained by taking averages of direction-dependent tensors at Λ over the 2-sphere of the η's. We write this operation "Av." It is not difficult to prove the following important relation connecting averages and derivatives:

$$\mathrm{Av}\,\partial_m \mathbf{T}^{a\cdots c}{}_{b\cdots d} = \mathrm{Av}\,2\eta_m \mathbf{T}^{a\cdots c}{}_{b\cdots d}, \tag{27}$$

for any direction-dependent tensor $\mathbf{T}^{a\cdots c}{}_{b\cdots d}(\eta)$ at Λ.

As an example, we exhibit one conserved quantity in electrodynamics—the charge. Set

$$Q = \mathrm{Av}(\mathbf{E}_a \eta^a). \tag{28}$$

Then, from (26) and (27),

$$\begin{aligned} Q &= \mathrm{Av}[\boldsymbol{\epsilon}_{abc}\eta^a\partial^b(\boldsymbol{\varphi}\mathbf{B}^c)] \\ &= \mathrm{Av}[\partial^b(\boldsymbol{\epsilon}_{abc}\eta^a\boldsymbol{\varphi}\mathbf{B}^c) - \boldsymbol{\epsilon}_{abc}(\partial^b\eta^a)\boldsymbol{\varphi}\mathbf{B}^c] \\ &= \mathrm{Av}[(2\eta^b)\boldsymbol{\epsilon}_{abc}\eta^a\boldsymbol{\varphi}\mathbf{B}^c] \\ &= 0. \end{aligned} \tag{29}$$

Thus, Q is invariant under boosts, while its lack of indices and lack of dependence on η ensures invariance under rotations. Every asymptotic quantity is invariant under translations. This Q is simply a real number associated with any asymptotically well-behaved solution of Maxwell's equations.

3. ASYMPTOTIC BEHAVIOR IN CURVED SPACE

In this section we discuss the notion of asymptotic flatness of initial data for Einstein's equation. The treatment is essentially that of Sec. 2, modified to take account of the fact that the initial spacelike 3-surface S need not have a flat intrinsic metric. By means of a conformal transformation, one introduces a single point Λ at spatial infinity. One then introduces fields at the point Λ [analogous to (21)] and their constraint and evolution equations [analogous to (22) and (26)].

Let M be a four-dimensional manifold with a smooth metric g_{ab} of signature $(-,+,+,+)$. Let S be a three-dimensional, spacelike submanifold of M. Denote by h_{ab} the (positive-definite) induced metric on S, and by p^{ab} the extrinsic curvature[15] of S (see Ref. 9 or Appendix A). These tensor fields on S describe the intrinsic geometry of S and the embedding of S in M, respectively. Suppose that our space–time satisfies Einstein's equation without sources, i.e., that the Ricci tensor[16] R_{ab} of (M, g_{ab}) vanishes. It then follows (see Ref. 9 or Appendix A) that h_{ab} and p^{ab} satisfy the constraint equations:

$$\mathcal{R} - p_{mn}p^{mn} + p^2 = 0, \tag{30}$$

$$D_m(p^{am} - p h^{am}) = 0, \tag{31}$$

where D_a is the (covariant) derivative on S with respect to h_{ab}, $\mathcal{R}$ is the scalar curvature[16] of S, and $p = p^m_m$. (Indices of tensor fields on S are raised and lowered with the metric h_{ab} and its inverse.) More generally, a 3-manifold S with a pair of smooth symmetric tensor fields h_{ab} (positive-definite) and p^{ab}, subject to (30) and (31), will be called an *initial-data set*. Our description of the gravitational field at spatial infinity will be in terms of initial-data sets.

The evolution equations determine the data at a slightly different "time," i.e., on a nearby surface, in terms of the data on S. Let φ be an arbitrary (not necessarily positive) smooth function on S. For fixed constant ϵ, we denote by S' the surface which results from moving each point x of S a proper distance $\epsilon\varphi(x)$ along the future-directed geodesic normal to S through x. This construction clearly gives a natural mapping from S' to S. Hence, the initial data on S' defines a second set $h_{ab}(\epsilon)$, $p^{ab}(\epsilon)$, of initial data on S. Setting $\dot{h}_{ab} = (d/d\epsilon)h_{ab}(\epsilon)|_{\epsilon=0}$, $\dot{p}^{ab} = (d/d\epsilon)p^{ab}(\epsilon)|_{\epsilon=0}$, we have the *evolution equations* (see Ref. 9 or Appendix A):

$$\dot{h}_{ab} = 2\varphi p_{ab}, \tag{32}$$

$$\dot{p}^{ab} = D^aD^b\varphi - \varphi\mathcal{R}^{ab} - 2\varphi p^{am}p^b{}_m - \varphi p p^{ab}, \tag{33}$$

where $\mathcal{R}^{ab}$ is the Ricci tensor[16] of S. (Note the placement of indices: $\dot{p}_{ab} = (h_{am}h_{bn}p^{mn})^{\cdot} \neq h_{am}h_{bn}\dot{p}^{mn}$; $\dot{h}^{ab} = -2\varphi p^{ab} \neq h^{am}h^{bn}\dot{h}_{mn}$.) The function φ in (32) and (33) allows the evolution to proceed at different rates at different points of S, and even into the future at certain points of S (where $\varphi > 0$) and into the past at others ($\varphi < 0$). The constraints are preserved under the evolution, for, applying a dot to both sides of (30) and (31), and using (32) and (33), we obtain identities (see Appendix B).

We shall be concerned only with the asymptotic behavior of an initial-data set. The discussion of Sec. 2 suggests that, in order to describe this asymptotic behavior, we apply a conformal transformation to Eqs. (30)–(33). Set

$$\tilde{h}_{ab} = \Omega^2 h_{ab}, \tag{34}$$

where Ω is a positive scalar field on S. As in Sec. 2, we assign to φ dimensions[12] sec:

$$\tilde{\varphi} = \Omega\varphi. \tag{35}$$

Finally, it is convenient to choose for p^{ab} dimensions[12] sec^{-1}:

$$\tilde{p}^{ab} = \Omega^{-3}p^{ab}, \quad \tilde{p}^a{}_b = \Omega^{-1}p^a{}_b, \quad \tilde{p}_{ab} = \Omega p_{ab}. \tag{36}$$

Substituting (34)–(36) into (30)–(33), we obtain (Appendix B)

$$\tilde{\mathcal{R}} - \tilde{p}^{ab}\tilde{p}_{ab} + \tilde{p}^2 = 6\Omega^{-2}(\tilde{D}^m\Omega)(\tilde{D}_m\Omega) - 4\Omega^{-1}\tilde{D}^2\Omega, \tag{37}$$

$$\tilde{D}_m(\tilde{p}^{am} - \tilde{p}\tilde{h}^{am}) = 2\Omega^{-1}(\tilde{D}_m\Omega)\tilde{p}^{am}, \tag{38}$$

$$\dot{\tilde{h}}_{ab} = 2\tilde{\varphi}\tilde{p}_{ab} + 2\Omega^{-1}\dot{\Omega}\tilde{h}_{ab}, \tag{39}$$

$$\begin{aligned}\dot{\tilde{p}}^{ab} = &-3\Omega^{-1}\dot{\Omega}\tilde{p}^{ab} + \tilde{D}^a\tilde{D}^b\tilde{\varphi} - \tilde{\varphi}\tilde{\mathcal{R}}^{ab} - 2\tilde{\varphi}\tilde{p}^{am}\tilde{p}^b{}_m \\ &- \tilde{\varphi}\tilde{p}\tilde{p}^{ab} - 2\Omega^{-1}\tilde{\varphi}\tilde{D}^a\tilde{D}^b\Omega + 3\Omega^{-2}\tilde{h}^{ab}(\tilde{D}^m\Omega)(\tilde{D}_m\Omega)\tilde{\varphi} \\ &- \Omega^{-1}\tilde{h}^{ab}(\tilde{D}^m\tilde{\varphi})(\tilde{D}_m\Omega) - \Omega^{-1}\tilde{\varphi}\tilde{h}^{ab}\tilde{D}^2\Omega,\end{aligned} \tag{40}$$

where $\tilde{D}_a$, $\tilde{\mathcal{R}}^{ab}$, and $\tilde{\mathcal{R}}$ are the derivative, Ricci tensor, and scalar curvature, respectively, with respect to $\tilde{h}_{ab}$, and where indices of tensor fields with a tilde are raised and lowered with $\tilde{h}_{ab}$ and its inverse. Note that Ω appears in (39) and (40): We permit the conformal factor to vary in an essentially arbitrary way from one surface to the next. In other words, we include the conformal behavior of the fields in the evolution equations.

We now introduce the notion of asymptotic flatness of an initial-data set. We first require some general conditions on the metric h_{ab} which reflect the intuitive idea that "h_{ab} approaches a Euclidean metric sufficiently rapidly at infinity." (The situation is essentially that of Ref. 7.) We require, firstly, that it be possible to attach a single point Λ to S to obtain a second 3-manifold $\tilde{S} = S \cup \Lambda$. Secondly, we require that it be possible to assign to $\tilde{S}$ a metric $\tilde{h}_{ab}$ which is C^0 at Λ and C^∞ elswhere, and a scalar field Ω which is C^2 at Λ and C^∞ elsewhere. Finally, we require that on $S = \tilde{S} - \Lambda$, $\tilde{h}_{ab} = \Omega^2 h_{ab}$, and that, at the point Λ,

$$\Omega = 0, \quad \tilde{D}_a\Omega = 0, \quad \tilde{D}_a\tilde{D}_b\Omega = 2\tilde{h}_{ab}. \tag{41}$$

(Note that Eqs. (41) do not require a connection.)

These conditions may be compared with the discussion preceeding Eq. (11). Asymptotic flatness of a 3-space S, h_{ab} in the curved case is defined to mean the existence of a conformal completion by a point at infinity which shares certain properties with the standard conformal completion of Euclidean 3-space. The essential difference between the flat and curved cases involves the choice of differentiability conditions. (Compare, Ref. 7.) The particular choice of differentiability conditions above is motivated by the fact (which we shall see in more detail later) that one cannot impose stronger conditions even in very simple examples (e.g., in the Weyl solutions). Finally, we remark that, by a generalization of Ref. 7, it can be shown that the completion above is unique if it exists.

The conditions above do not suffice, however, to define a reasonable notion of asymptotic flatness of an initial-data set. Roughly speaking, the conditions, above are on the *geometry:* We require further conditions, analogous to (21), on the *fields*. We use the terminology and notation of Sec. 2.[17] [See Eqs. (16)–(20).] The required conditions are that

$$\mathbf{p}^{ab}(\eta) = \lim \tilde{p}^{ab}, \tag{42}$$

$$\mathfrak{R}_{ab}(\eta) = \lim(\Omega^{1/2})\tilde{\mathfrak{R}}_{ab}, \tag{43}$$

$$\mathbf{\Omega}_{ab}(\eta) = \lim \Omega^{-1/2}(\tilde{D}_a\tilde{D}_b\Omega - 2\tilde{h}_{ab}) \tag{44}$$

define direction-dependent tensors at Λ.

To summarize, an initial-data set S, h_{ab}, p^{ab} will be said to be *asymptotically flat* if there exists a 3-manifold $\tilde{S} = S \cup \Lambda$ with a positive-definite metric $\tilde{h}_{ab}$ and a nonnegative scalar field Ω such that:

(1) At Λ, $\tilde{h}_{ab}$ is C^0 and Ω is C^2.

(2) On S, $\tilde{h}_{ab} = \Omega^2 h_{ab}$.

(3) At Λ, $\Omega = 0$, $\tilde{D}_a\Omega = 0$, and $\tilde{D}_a\tilde{D}_b\Omega = 2\tilde{h}_{ab}$.

(4) Equations (42)–(44) define direction-dependent tensors at Λ.

On what basis have these particular conditions been selected? An appropriate definition of asymptotic flatness of initial-data sets should satisfy one important criterion: It should be true that, if an initial-data set is asymptotically flat, and if that set is evolved [using (32) and (33) for an appropriate class of φ's], then the resulting initial data sets are also asymptotically flat. Suppose, for example, that we were to omit (42) from the definition of asymptotic flatness. Then, in general, evolution under (32) and (33) would destroy (43) and (44). If, on the other hand, (42) were replaced by a stronger condition such as $\mathbf{p}^{ab} = 0$, then, even in static space–times, this stronger condition would be destroyed by (32) and (33). Roughly speaking, the conditions above are the most stringent which are still satisfied by the static space–times.[7] [Conditions (42)–(45) are essentially those of ADM.[5]]

The asymptotic gravitational field will be characterized by the three direction-dependent tensors $\mathbf{p}^{ab}(\eta)$, $\mathfrak{R}_{ab}(\eta)$, and $\mathbf{\Omega}_{ab}(\eta)$ at Λ. These are analogous to the $\mathbf{E}_a(\eta)$ and $\mathbf{B}_a(\eta)$ of Sec. 2. We now derive the constraint equations, the gravitational analogs of (22). As in Sec. 2, we denote[18] by $\mathbf{h}_{ab}$ the direction-dependent tensor at Λ defined by $\tilde{h}_{ab}$, and we use this metric to raise and lower indices of direction-dependent tensors at Λ. Taking the limit of (38) at Λ and using (17), we obtain immediately

$$\partial_m(\mathbf{p}^{am} - \mathbf{p}\mathbf{h}^{am}) = 4\eta_m \mathbf{p}^{am}. \tag{45}$$

Multiply Eq. (37) by $(\Omega)^{1/2}$:

$$\Omega^{1/2}(\tilde{\mathfrak{R}} - \tilde{p}^{ab}\tilde{p}_{ab} + \tilde{p}^2) = \Omega^{1/2}(6\Omega^{-2}(\tilde{D}_m\Omega)(\tilde{D}^m\Omega) - 4\Omega^{-1}\tilde{D}^2\Omega). \tag{46}$$

It seems at first that Eq. (46) will lead to difficulties, for whereas the left side of this equation certainly defines a direction-dependent tensor at Λ, the right side appears to be unbounded at Λ. However, it can be checked, using (41), that the two terms on the right in (46) (both unbounded) just cancel out to highest order. By taking the limit of (46) at Λ and using (42)–(44), we obtain the second constrain equation

$$\mathfrak{R} = -4\mathbf{\Omega}^m{}_m + 8\eta^m\eta^n\mathbf{\Omega}_{mn}. \tag{47}$$

Note that the terms involving $\tilde{p}^{ab}$ in (46) do not contribute to (47).

There are two additional constraint equations which arise directly from the definitions (43) and (44). By multiplying the Bianchi identity on $\tilde{\mathfrak{R}}_{ab}$,

$$\tilde{D}_m(\tilde{\mathfrak{R}}^{am} - \tfrac{1}{2}\tilde{\mathfrak{R}}\tilde{h}^{am}) = 0, \tag{48}$$

by Ω and by taking the limit at Λ using (43), we obtain

$$(\partial^m - \eta^m)(\mathfrak{R}_{am} - \mathfrak{R}\mathbf{h}_{am}) = 0. \tag{49}$$

Similarly, by taking the limit of the identity

$$\tilde{D}_{[a}\tilde{D}_{b]}\tilde{D}_c\Omega = \tfrac{1}{2}\tilde{\mathfrak{R}}_{abcd}\tilde{D}^d\Omega \tag{50}$$

at Λ using (43) and (44) we have

$$\partial_{[a}\mathbf{\Omega}_{b]c} - \eta_{[a}\mathbf{\Omega}_{b]c} = \mathfrak{R}_{abcd}\eta^d, \tag{51}$$

where we have defined[19]

$$\mathfrak{R}_{abcd} = 2\mathbf{h}_{a[c}(\mathfrak{R}_{d]b} - \tfrac{1}{4}\mathfrak{R}\mathbf{h}_{d]b}) - 2\mathbf{h}_{b[c}(\mathfrak{R}_{d]a} - \tfrac{1}{4}\mathfrak{R}\mathbf{h}_{d]a}). \tag{52}$$

Thus, the asymptotic gravitational field is described by the direction-dependent tensors $\mathbf{p}^{ab}$, $\mathfrak{R}_{ab}$, and $\mathbf{\Omega}_{ab}$ at Λ, subject to (45), (47), (49), and (51).

In order to derive the evolution equations, we must fix the asymptotic behavior of the fields $\tilde{\varphi}$ and $\dot{\Omega}$ in more detail. Consider first the function $\tilde{\varphi}$, which characterizes the rate of evolution at each point of S. In the flat case, Eqs. (23) are the asymptotic conditions on $\tilde{\varphi}$ which reflect the fact that this function represents an infinitesimal Poincaré transformation In order that $\tilde{\varphi}$ represent an "asymptotic Poincaré transformation" in the curved case, we wish to impose as many of these conditions as possible. We can certainly require that $\tilde{\varphi}$ vanish at Λ, and that the derivative of $\tilde{\varphi}$, $s_a = \tilde{D}_a\tilde{\varphi}$ exist at Λ. That is, we can require that (24) define a direction-dependent scalar $\boldsymbol{\varphi}(\eta)$ at Λ which is linear, i.e., which satisfies

$$\partial_{(a}\partial_{b)}\boldsymbol{\varphi} = -\boldsymbol{\varphi}\mathbf{h}_{ab} + \boldsymbol{\varphi}\eta_a\eta_b - \eta_{(a}\partial_{b)}\boldsymbol{\varphi}. \tag{53}$$

There is, however, a difficulty associated with the third condition (23). Define the direction-dependent tensor

$$\boldsymbol{\varphi}_{ab} = \lim \tilde{D}_a\tilde{D}_b\tilde{\varphi} \tag{54}$$

at Λ. Then the identity $\tilde{D}_{[a}\tilde{D}_{b]}\tilde{D}_c\tilde{\varphi} = \tfrac{1}{2}\tilde{\mathfrak{R}}_{abcd}\tilde{D}^d\tilde{\varphi}$ implies

$$\partial_{[a}\boldsymbol{\varphi}_{b]c} = \tfrac{1}{2}\mathfrak{R}_{abcd}(\partial^d\boldsymbol{\varphi} + \eta^d\boldsymbol{\varphi}). \tag{55}$$

The third condition (23), expressed in terms of $\boldsymbol{\varphi}_{ab}$, is $\boldsymbol{\varphi}_{ab} = 2s\mathbf{h}_{ab}$, where s is a constant. But this is inconsistent with (55). That is, the metric in the curved case does not approach a Euclidean metric at infinity sufficiently rapidly that temporal translations can be clearly distinguished in the presence of boosts. This problem will be discussed in more detail in the following section, but for the present we shall merely regard the asymptotic behavior of $\tilde{\varphi}$ as described by the direction-dependent tensors $\boldsymbol{\varphi}(\eta)$ and $\boldsymbol{\varphi}_{ab}(\eta)$ at Λ, subject to (53) and (55).

We next consider the asymptotic conditions on $\dot{\Omega}$, the quantity which describes an infinitesimal conformal transformation. By taking the limit of (39) at Λ and using the vanishing of $\tilde{\varphi}$ there and (42), we have

$$\dot{\tilde{h}}_{ab} = 2\Omega^{-1}\dot{\Omega}\tilde{h}_{ab}. \tag{56}$$

On the other hand, in order that (41) be preserved under the conformal transformation, we must have

$$\tilde{D}_a\tilde{D}_b\dot{\Omega} = 2\dot{\tilde{h}}_{ab} \tag{57}$$

at Λ. Equations (56) and (57) together imply that $\Omega^{-1}\dot{\Omega}$ vanishes at Λ. Hence, we require the vanishing of $\Omega^{-1}\dot{\Omega}$ at Λ in order to preserve asymptotic flatness. But now Eq. (56) implies that $\dot{\tilde{h}}_{ab}$ vanishes at Λ. That is to say, $\mathbf{h}_{ab}$ remains fixed throughout the course of the evolution. This is an important conclusion, for much of the algebraic structure at Λ is determined by $\mathbf{h}_{ab}$. The effect of $\dot{\Omega}$ on the evolution can be expressed in terms of the direction-dependent scalar

$$\omega(\eta) = \lim(\Omega^{-3/2})\dot{\Omega}. \tag{58}$$

Thus, the evolution is described by three direction-dependent tensors at Λ as $\varphi(\eta)$ (boosts), $\varphi_{ab}(\eta)$ (generalized time translations), and $\omega(\eta)$ (generalized space translations), subject to (53) and (55). (Rotations, of course, are represented by ordinary rotations in the tangent space at Λ.) In the flat case, we can require that φ_{ab} be a multiple of $\mathbf{h}_{ab}$ and that $\omega(\eta)$ be linear in η, thus recovering the Poincaré group.

We now write down the evolution equations. By evaluating the rate of change of the Ricci tensor from (39), and taking the limit at Λ, we obtain

$$\begin{aligned}\dot{\mathfrak{R}}_{ab} = &-\partial_{(a}\partial_{b)}\omega - \eta_{(a}\partial_{b)}\omega - 3\omega\mathbf{h}_{ab} + \omega\eta_a\eta_b - \mathbf{h}_{ab}\partial^2\omega\\ &+ \varphi\partial_{(a}\partial_{b)}\mathbf{p} - \varphi\partial^2\mathbf{p}_{ab} + 10\varphi\eta^m\partial_{(a}\mathbf{p}_{b)m} - \varphi\eta_{(a}\partial_{b)}\mathbf{p}\\ &- 2(\partial^m\varphi)\partial_m\mathbf{p}_{ab} + 2(\partial^m\varphi)\partial_{(a}\mathbf{p}_{b)m} + 8\eta^m\mathbf{p}_{m(a}\partial_{b)}\varphi\\ &- 8\varphi\eta_{(a}\mathbf{p}_{b)m}\eta^m + 8\varphi\mathbf{p}_{ab}.\end{aligned} \tag{59}$$

By taking the limit of (40) at Λ,

$$\begin{aligned}\mathbf{p}_{ab} = \varphi_{ab} &- \mathbf{h}_{ab}\eta^m\eta^n\varphi_{mn} - \varphi\mathfrak{R}_{ab} - 2\varphi\Omega_{ab} - \varphi\mathbf{h}_{ab}\Omega_m^m\\ &+ \tfrac{7}{2}\varphi\mathbf{h}_{ab}\eta^m\eta^n\mathbf{\Omega}_{mn} - \tfrac{1}{2}\mathbf{h}_{ab}\mathbf{\Omega}_{mn}\eta^m\partial^n\varphi.\end{aligned} \tag{60}$$

Finally, from (44), (56), and (58), we obtain

$$\begin{aligned}\dot{\mathbf{\Omega}}_{ab} = &\partial_{(a}\partial_{b)}\omega + \eta_{(a}\partial_{b)}\omega + \omega\mathbf{h}_{ab} - \omega\eta_a\eta_b\\ &- 2\varphi\mathbf{p}_{ab} - 4\varphi\eta_{(a}\mathbf{p}_{b)m}\eta^m - 4\eta^m\mathbf{p}_{m(a}\partial_{b)}\varphi - 4\varphi\eta^m\mathbf{p}_{m(a}\eta_{b)}.\end{aligned} \tag{61}$$

These three equations give the values of the fields $\mathbf{p}^{ab}$, $\mathfrak{R}_{ab}$, and $\mathbf{\Omega}_{ab}$ associated with a nearby surface S' in terms of the values associated with S and the fields φ, φ_{ab}, and ω, which describe the asymptotic behavior of S' relative to S. It can be checked directly (although the calculation is rather long) that Eqs. (59)–(61) preserve the constraint equations (45), (47), (49), and (51).

4. ASYMPTOTIC SYMMETRIES

Consider an asymptotically flat initial-data set. The asymptotic gravitational field is described by the three direction-dependent tensors $\mathfrak{R}_{ab}$, $\mathbf{p}_{ab}$, and Ω_{ab}, subject to (45), (47), (49), and (51). The evolution (i.e., the effect of an infinitesimal change in the surface S and in the conformal factor) is described by the direction-dependent tensors φ, φ_{ab}, and ω, subject to (53) and (55). The asymptotic gravitational field evolves according to (59)–(61).

In the electromagnetic case, by contrast, the asymptotic field is given by the direction-dependent vectors $\mathbf{E}_a$ and $\mathbf{B}_a$ subject to (22). The evolution is described by the direction-dependent scalar φ, subject to (25). The evolution equations are (26).

One significant difference between these two cases is that, whereas the single quantity φ suffices to determine the evolution in the (flat) electromagnetic case, three quantities φ, φ_{ab}, and ω are required in the gravitational case. What are the geometrical interpretations of these three quantities?

The direction-dependent scalar $\varphi(\eta)$ of Sec. 3 is completely analogous to the $\varphi(\eta)$ of Sec. 2. In both cases this scalar is linear in η[i.e., it satisfies (53)], and in both cases it represents an asymptotic boost.

Consider next the direction-dependent tensor φ_{ab} which enters (60). Comparing (54) and (23), we see that, in the flat case $\varphi_{ab} = 2s\mathbf{h}_{ab}$, where s is a constant. But, as we saw in Sec. 3, this choice for φ_{ab} is unacceptable in the curved case because of (55). Thus, whereas one has a one-dimensional family of temporal translations (labeled by s) in flat space, in the presence of curvature one is confronted by an infinite-dimensional family of "generalized temporal translations" (labeled by φ_{ab}).[20] [The general solution of (55) involves one arbitrary function.] This phenomenon is essentially the same as that which gives rise to supertranslations[2] in the study of asymptotic structure at null infinity. As discussed in Sec. 1, an enlargement of the translation subgroup is to be expected in the presence of curvature. Finally, note that φ_{ab} appears in (60) only in the combination $\varphi_{ab} - \mathbf{h}_{ab} \times \eta^m\eta^n\varphi_{mn}$. Thus, if it were possible to set $\varphi_{ab} = 2s\mathbf{h}_{ab}$, the constant s would drop out of (60). This remark is the gravitational analog of the fact that s does not appear in the evolution equations for electrodynamics, Eq. (26).

Finally, we consider the direction-dependent scalar ω which appears in (59) and (61). The presence of ω reflects the fact that $\mathfrak{R}_{ab}$ and Ω_{ab} are not conformally invariant. In the electromagnetic case, on the other hand, $\mathbf{E}_a$ and $\mathbf{B}_a$ are conformally invariant. The reason for this difference is that "conformally invariant" has different meanings in the two cases. For S Euclidean, one can choose the metric $\tilde{h}_{ab}$ to be smooth at Λ. One can then require that a conformal transformation preserve this smoothness, i.e., one can admit only conformal factors of the form[21] $(1 + \Omega^{1/2}\omega)$ near Λ, where ω is linear in η. In the curved case, on the other hand, the metric $\tilde{h}_{ab}$ is only continuous at Λ. The conformal factor $(1 + \Omega^{1/2}\omega)$ near Λ preserves continuity for any direction-dependent scalar ω at Λ. Thus, there is no criterion, in the presence of curvature, for selecting the three-dimensional class of admissible ω's from what is, *a priori*, an infinite-dimensional class.[22] As we saw in Sec. 2, conformal transformations near Λ are associated with spatial translations. Curvature has the effect of replacing the usual three-dimensional family of ordinary spatial translations by an infinite-dimensional class of generalized spatial translations. Finally, note that ω appears in (59) and (61) only in the combination $\partial_{(a}\partial_{b)}\omega + \eta_{(a}\partial_{b)}\omega + \omega(\mathbf{h}_{ab} - \eta_a\eta_b)$. Thus, if ω were linear in η, it would drop out of (59) and (61). This remark is the gravitational analog of conformal invariance in the electromagnetic case.

One could imagine several courses of action in response to the appearance of generalized translations.

One could simply accept as the asymptotic symmetry group at spatial infinity the infinite-dimensional Lie group consisting of the boosts, rotations (i.e., rotations in the tangent space at Λ), and generalized translations. This group has essentially[23] the same structure as the Bondi–Metzner–Sachs group.[2] One could regard all of the objects $\mathcal{R}_{ab}$, $\mathbf{p}_{ab}$, and Ω_{ab} as having physical significance, and accept (59)–(61) as the final evolution equations. The difficulty with this approach is that it apparently precludes, as we shall see in the following section, the introduction of a finite-dimensional surface consisting of all points at spatial infinity. The existence of such a surface has been particularly useful (e.g., as a domain of integration) in the null case,[4] and so it would be unfortunate if no finite-dimensional surface could be obtained in the spatial case. Furthermore, an infinite-dimensional asymptotic symmetry groups leads to difficulties in the interpretation of asymptotic quantities.

Alternatively, one might separate a family of "ordinary" temporal and spatial translations from the much larger class of generalized translations by means of subsidiary conditions. Suppose, for example, that we impose, as additional conditions on the asymptotic fields, the following:

$$\mathbf{p} = 0, \quad \mathcal{R} = 0. \tag{62}$$

It is not difficult to check, from (59) and (60), that there always exists a choice of φ_{ab} and ω for which (62) is preserved under the evolution, and that, furthermore, ω is thus determined uniquely up to the addition of a scalar linear in η, while φ_{ab} is determined uniquely up to a constant multiple of $\mathbf{h}_{ab}$. The subsidiary conditions (62) thus restore the Poincaré group as the asymptotic symmetry group. The problem with this approach is that there are, presumably, other equally reasonable choices of subsidiary conditions besides (62). Hence, it would be difficult to interpret physically conclusions drawn from this approach because it would be difficult to ascribe physical significance to conditions such as (62).

It turns out, fortunately, that a great deal of the structure of the asymptotic gravitational field can be discussed without any commitment whatever as to how the generalized translations will be treated. We ask if it is possible to obtain direction-dependent tensors which are expressed in terms of $\mathcal{R}_{ab}$, $\mathbf{p}_{ab}$, and Ω_{ab}, but whose evolution involves neither φ_{ab} nor ω. It turns out that, not only do there exist such tensors, but they carry most of the information carried by $\mathcal{R}_{ab}$, $\mathbf{p}_{ab}$, and Ω_{ab}. Furthermore, the information these tensors do carry appears to be the most interesting from a physical viewpoint. What remains is dominated by the generalized translations, in the sense that its evolution can be given arbitrarily by appropriate choices of φ_{ab} and ω. By restricting attention to these new tensors, the generalized translations drop out of the formalism. Only the Lorentz group survives.

Consider first the appearance of $\boldsymbol{\omega}$ in (59) and (61). It follows from (61) that the quantity

$$\boldsymbol{\Omega}_a = \boldsymbol{\Omega}_{ab}\eta^b \tag{63}$$

evolves independently of $\boldsymbol{\omega}$. Similarly, from (59) and (61), the combination

$$\mathcal{R}'_{ab} = \mathcal{R}_{ab} + \Omega_{ab} + \mathbf{h}_{ab}\boldsymbol{\Omega}^m{}_m \tag{64}$$

evolves independently of ω. Consider next the appearance of φ_{ab} in (60). Clearly, the quantity

$$\beta = -3\mathbf{p}_{mn}\eta^m\eta^n \tag{65}$$

evolves independently of φ_{ab}. Furthermore, it can be checked from (55) and (60) that the combination

$$\mathbf{K}_{abc} = \partial_{[a}\mathbf{p}_{b]c} + 2\eta^m\mathbf{p}_{m[a}\mathbf{h}_{b]c} \tag{66}$$

also evolves independently of $\boldsymbol{\varphi}_{ab}$. It is not enough, however, merely to find quantities $\boldsymbol{\Omega}_a$, $\mathcal{R}'_{ab}$, β, and $\mathbf{K}_{abc}$, whose evolution is independent of $\boldsymbol{\varphi}_{ab}$ and ω. It must also be true that the evolution of $\boldsymbol{\Omega}_a$, $\mathcal{R}'_{ab}$, β, and $\mathbf{K}_{abc}$ depends only on the values of these four quantities and not on the parts of $\mathcal{R}_{ab}$, $\mathbf{p}_{ab}$, and $\boldsymbol{\Omega}_{ab}$, which have been discarded in (63)–(66). This turns out to be the case. Although these new evolution equations can be derived from (59)–(61), we shall not write them here, as they are rather complicated and will not be needed.

Of course, the new fields $\boldsymbol{\Omega}_a$, $\mathcal{R}'_{ab}$, β, and $\mathbf{K}_{abc}$ satisfy certain constraint equations. These are obtained from the constraint equations (45), (47), (49), and (51) on our original fields:

$$\partial_a\boldsymbol{\Omega}_b = 2(\mathbf{h}_a{}^m - \eta_a\eta^m)(\mathbf{h}_b{}^n - \eta_b\eta^n) \times (\mathcal{R}'_{mn} - \tfrac{1}{2}\mathcal{R}'\mathbf{h}_{mn}) - 2\mathbf{h}_{ab}(\boldsymbol{\Omega}_m\eta^m) + 2\boldsymbol{\Omega}_a\eta_b, \tag{67}$$

$$\mathcal{R}'^m{}_m = 8\boldsymbol{\Omega}_m\eta^m, \tag{68}$$

$$\partial^m(\mathcal{R}'_{am} - \tfrac{1}{2}\mathcal{R}'\mathbf{h}_{am}) = 3\mathcal{R}'_{am}\eta^m - \tfrac{1}{2}\mathcal{R}'\eta_a - 4\boldsymbol{\Omega}_a, \tag{69}$$

$$\mathbf{K}_{[abc]} = 0, \quad \mathbf{K}_{am}{}^m = 0, \tag{70}$$

$$\mathbf{K}_{amn}\eta^m\eta^n = -\tfrac{1}{6}\partial_a\beta, \tag{71}$$

$$\partial_{[a}\mathbf{K}_{bc]d} = 2\mathbf{h}_{d[a}\mathbf{K}_{bc]m}\eta^m + \eta_{[a}\mathbf{K}_{bc]d}. \tag{72}$$

It is natural to ask at this point whether or not it is possible to eliminate some or all of (67)–(72). Specifically, can the general solution of (67)–(72) be expressed in terms of some collection of essentially arbitrary fields? It turns out that such potentials do indeed exist. Their introduction enormously simplifies all our equations.

We begin with the "geometrical" fields $\boldsymbol{\Omega}_a$ and $\mathcal{R}'_{ab}$. Set

$$\boldsymbol{\alpha} = \tfrac{1}{2}\boldsymbol{\Omega}_a\eta^a. \tag{73}$$

Then, by contracting (67) with η^b, we have

$$\boldsymbol{\Omega}_a = \tfrac{2}{3}\partial_a\boldsymbol{\alpha} + 2\boldsymbol{\alpha}\eta_a. \tag{74}$$

Via substituting (74) into (67) and using (68),

$$\mathcal{R}'_{ab} = \tfrac{1}{3}\partial_{(a}\partial_{b)}\boldsymbol{\alpha} + \tfrac{1}{3}\eta_{(a}\partial_{b)}\boldsymbol{\alpha} + 11\boldsymbol{\alpha}(\mathbf{h}_{ab} - \eta_a\eta_b) + (\tfrac{1}{3}\partial^2\boldsymbol{\alpha} + 6\boldsymbol{\alpha}) \times (2\eta_a\eta_b - \mathbf{h}_{ab}) + 2\eta_{(a}\mathcal{R}'_{b)}, \tag{75}$$

where $\mathcal{R}'_a = (\mathbf{h}_a{}^m - \eta_a\eta^m)\eta^n\mathcal{R}'_{mn}$. Substituting (75), Eq. (69) reduces to

$$\partial^m\mathcal{R}'_m = 0. \tag{76}$$

But (76) and the fact that $\mathcal{R}'_m \eta^m = 0$ together imply

$$\partial_{[a}(\epsilon_{b]cd}\, \eta^c\, \mathcal{R}'^d) = \eta_{[a}(\epsilon_{b]cd}\, \eta^c\, \mathcal{R}'^d). \tag{77}$$

But Eq. (77) is the integrability condition for $\epsilon_{bcd}\eta^c \mathcal{R}'^d$ to be the ∂ derivative of some direction-dependent scalar. [See (20).] Hence,

$$\mathcal{R}'_a = -4\epsilon_{abc}\, \eta^b \partial^c \delta \tag{78}$$

for some $\delta(\eta)$ (unique up to a constant). Thus, the most general solution $\Omega_a, \mathcal{R}'_{ab}$ of Eqs. (67)–(69) can be expressed, via (74), (75), and (78), in terms of a direction-dependent scalar α and a second direction-dependent scalar δ determined only up to a constant.

A similar, but slightly more complicated, reduction is possible for $\mathbf{K}_{abc}$. From (70) to (71),

$$\mathbf{K}_{abc} = \tfrac{2}{3}\eta_c \eta_{[a}\, \partial_{b]} \beta + \tfrac{1}{3}\mathbf{h}_{c[b}\, \partial_{a]}\beta + \tfrac{1}{2}\eta_c \mathbf{D}\, \epsilon_{abd}\, \eta^d - \tfrac{1}{2}\mathbf{D}\, \eta_{[a}\, \epsilon_{b]cd}\, \eta^d + 2\eta_{[a}\, \mathbf{E}_{b]m}\, \epsilon^m{}_{cn}\, \eta^n, \tag{79}$$

where $\mathbf{E}_{ab}$ satisfies

$$\mathbf{E}_{ab} = \mathbf{E}_{(ab)}, \quad \mathbf{E}^m{}_m = 0, \quad \eta^m \mathbf{E}_{am} = 0, \tag{80}$$

and where $\mathbf{D} = \eta_a \epsilon^{abc} \mathbf{K}_{bcd}\, \eta^d$. By substituting (79) into (72) and using (80),

$$\partial^m \mathbf{E}_{am} = -\tfrac{1}{4}\partial_a \mathbf{D}. \tag{81}$$

Set

$$\partial^2 \gamma + 2\gamma = \tfrac{1}{4}\mathbf{D}. \tag{82}$$

There always exists[24] a solution $\gamma(\eta)$ of (82), with γ determined by (82) uniquely up to the addition of a term linear in η. Now consider the combination $\mathbf{E}'_{ab} = \mathbf{E}_{ab} + 2\partial_{(a}\partial_{b)}\gamma + 2\eta_{(a}\partial_{b)}\gamma - \partial^2\gamma(\mathbf{h}_{ab} - \eta_a \eta_b)$. Then $\mathbf{E}'_{ab}$ is symmetric and trace-free, and has vanishing contractions with η. Furthermore, (81) and (82) imply $\partial^m \mathbf{E}'_{am} = 0$. But these conditions imply $\mathbf{E}'_{ab} = 0$.[25] Hence,

$$\mathbf{E}_{ab} = -2\partial_{(a}\partial_{b)}\gamma - 2\eta_{(a}\partial_{b)}\gamma + \partial^2\gamma(\mathbf{h}_{ab} - \eta_a \eta_b). \tag{83}$$

[Note that the addition to γ of a term linear in η does not affect the right side of (83).] Thus, the most general solution $\beta, \mathbf{K}_{abc}$ of (71)–(73) can be expressed, via (79), (82), and (83), in terms of the direction-dependent scalars β and γ, with γ determined only up to addition of a term linear in η.

The four direction-dependent tensors $\Omega_a, \mathcal{R}'_{ab}, \beta$, and $\mathbf{K}_{abc}$, subject to (67)–(72), have now been reduced to four direction-dependent scalars α, β, γ, and δ.

Finally, we obtain the evolution equations. By applying a dot to each of α, β, γ, and δ and using (59)–(61), we obtain

$$\dot{\alpha} = \varphi\beta, \tag{84}$$

$$\dot{\beta} = \partial^m[\varphi\, \partial_m \alpha] + 3\varphi\alpha, \tag{85}$$

$$\dot{\gamma} = \varphi\delta, \tag{86}$$

$$\dot{\delta} = \partial^m[\varphi\, \partial_m \gamma] + 3\varphi\gamma. \tag{87}$$

It is remarkable that the relatively complicated evolution and constraint equations of Sec. 3 should now become so simple. This feature presumably reflects the fact that, asymptotically, Einstein's equation reduces to the equation for a pure spin-2 field. Note that the addition of a linear term to γ on the right in (87) affects the evolution of δ by a constant term, and, conversely, the addition of a constant term to δ on the right in (86) affects the evolution of γ by a linear term.

To summarize, that part of the asymptotic gravitational field which is insensitive to generalized translations is described by four direction-dependent scalars α, β, γ, and δ, where the addition of a constant term to δ or a linear term to γ has no physical significance. These four scalars are acted upon by the Lorentz group via (84)–(87). Thus, we have a (infinite-dimensional) representation of the Lorentz group.

5. SPATIAL INFINITY

In the treatment of asymptotic structure at null infinity,[4] one introduces a three-dimensional surface $\mathcal{I}$ consisting of all points "at null infinity." The advantage of introducing such a surface is that, since asymptotic behavior is reformulated in terms of local behavior on $\mathcal{I}$, one can bring to bear the techniques of differential geometry on asymptotic problems. In this section, we shall obtain a similar surface in the spatial case.

We return for a moment to the flat case of Sec. 2. The asymptotic electromagnetic field is defined in terms of a spacelike three-dimensional plane S in Minkowski space. Since the asymptotic field is invariant under translations in Minkowski space, i.e., since it remains unchanged if S is replaced by any parallel plane, we can regard the asymptotic field as associated with an entire family of parallel planes. Let p denote a family of spacelike 3-planes in Minkowski space consisting of all planes parallel to a given plane. Let $\mathcal{S}$ denote the collection of all such families p of parallel planes. This $\mathcal{S}$ represents the set of points "at spatial infinity" for Minkowski space. Clearly, there is a natural, one-to-one correspondence between the collection $\mathcal{S}$ of families of parallel planes and the collection of all unit, future-directed, timelike vectors at a point O of Minkowski space. Thus, $\mathcal{S}$ has the structure of a three-dimensional manifold.

How does one interpret a vector in the manifold $\mathcal{S}$? Let p be the element of $\mathcal{S}$ consisting of all planes parallel to some plane S. A vector in $\mathcal{S}$ at p defines a nearby family p' of parallel planes, and so can be represented by an infinitesimal boost. But, as we saw in Sec. 2, the infinitesimal boosts relative to S, the constant vector fields in S, and the vectors s_a at the point Λ associated with S—three three-dimensional vector spaces—are all naturally isomorphic. Thus, vectors in the manifold $\mathcal{S}$ at the point p are in one-to-one correspondence with vectors at the point Λ defined by S. A similar correspondence therefore exists for tensors of arbitrary rank. In particular, the metric $\mathbf{h}_{ab}$ at Λ defines a positive-definite metric tensor field on $\mathcal{S}$. As one might expect, $\mathcal{S}$ has constant negative curvature with this metric. (The unit, future-directed timelike vectors at a point O of Minkowski space describe a hyperboloid.)

We next express the asymptotic fields $\mathbf{E}_a$ and $\mathbf{B}_a$ in terms of $\mathcal{S}$. For a fixed plane S, $\mathbf{E}_a(\eta)$ and $\mathbf{B}_a(\eta)$ are functions from unit vectors (η) at Λ to vectors at Λ. Thus, in terms of $\mathcal{S}$, $\mathbf{E}_a$ and $\mathbf{B}_a$ are to be regarded as mappings, for each point p of $\mathcal{S}$, from unit vectors at

p to vectors at p. In other words, $\mathbf{E}_a$ and $\mathbf{B}_a$ are practically vector fields on $\mathcal{S}$, except that the value of $\mathbf{E}_a$ and $\mathbf{B}_a$ depends not only on the choice of a point p of $\mathcal{S}$, but also on the choice of a unit vector in $\mathcal{S}$ at that point.[26]

The asymptotic symmetry group at spatial infinity in Minkowski space is just the group of isometries on $\mathcal{S}$, i.e., the Lorentz group.[27] Specifically, the action of the Lorentz group on $\mathcal{S}$ is as follows. A Lorentz transformation on Minkowski space takes a family p of parallel planes in Minkowski space to another family p'. That is to say, a Lorentz transformation takes each point of $\mathcal{S}$ to some other point of $\mathcal{S}$.

The presence of curvature requires minor modifications in the situation above. Instead of spacelike 3-planes in Minkowski space, one considers spacelike 3-surfaces in the space-time which are asymptotically flat (as initial-data sets). Let p denote a family of such surfaces, where any two surfaces in p differ by a generalized translation, and where any surface which can be obtained by a generalized translation from a surface in p is also in p. Denote by $\mathcal{S}$ the collection of all such families, so $\mathcal{S}$ is a three-dimensional manifold. Let the surface S in the space-time define the point p of $\mathcal{S}$. Then a vector s_a at the point Λ associated with S defines a nearby surface (up to generalized translations), and so defines a nearby point of $\mathcal{S}$. That is, vectors in $\mathcal{S}$ at the point p are again in one-to-one correspondence with vectors at the point Λ. So, $\mathcal{S}$ inherits a metric field from $\mathbf{h}_{ab}$ at Λ. As in the flat case, $\mathcal{S}$ has constant negative curvature, so the asymptotic symmetry group is again the (Lorentz) group of isometries on $\mathcal{S}$. Finally, the asymptotic gravitational field—i.e., the quantities $\boldsymbol{\alpha}, \beta, \gamma$, and δ,—represent real-valued functions on the collection of all unit vectors at points of $\mathcal{S}$.

Suppose one is given an asymptotically flat initial-data set. The discussion above suggests that, to obtain $\mathcal{S}$, one must solve (32) and (33) to obtain a four-dimensional space-time, determine the collection of all asymptotically flat 3-surfaces in this space-time, and group these surfaces into families consisting of surfaces which differ only by a generalized translation. In practice, however, this procedure can be avoided entirely. Choose a 3-manifold $\mathcal{S}$, topologically R^3, with a positive-definite metric, such that $\mathcal{S}$ is complete and has constant negative curvature. Choose any point p of $\mathcal{S}$ and any linear, metric-preserving mapping between vectors in $\mathcal{S}$ at p and vectors at Λ. (Since $\mathcal{S}$ is homogeneous and isotropic it makes no difference which point p or which mapping is chosen.) The $\boldsymbol{\alpha}, \beta, \gamma$, and δ associated with our original initial-data set thus become functions on unit vectors in $\mathcal{S}$ at p. A direction-dependent scalar $\varphi(\eta)$ at Λ, subject to (53), defines a vector s_a at Λ, and hence a vector in $\mathcal{S}$ at p. The evolution equations (84)–(87) can now be interpreted as giving the derivatives in $\mathcal{S}$ of $\boldsymbol{\alpha}, \beta, \gamma$, and $\boldsymbol{\delta}$. Evidently, we may iterate[28] (84)–(87) to define $\boldsymbol{\alpha}$, β, γ, and δ over all of $\mathcal{S}$. In this way, we obtain the manifold $\mathcal{S}$, its metric, and the four functions $\boldsymbol{\alpha}, \beta, \gamma$, and $\boldsymbol{\delta}$ on unit vectors in this manifold—all directly from a single asymptotically flat initial-data set. It is unnecessary to solve Einstein's equation. The reason for this possibility, of course, is that the asymptotic behavior at one time determines the behavior for all times.

We regard the three-dimensional manifold $\mathcal{S}$ in the spatial case as analogous to the three-dimensional manifold $\mathscr{I}$ in the null case. There is, however, one significant difference: whereas $\mathscr{I}$ is attached to the four-dimensional space-time as a boundary surface, $\mathcal{S}$ is not. Thus, for example, it is meaningful to ask for the endpoint on $\mathscr{I}$ of a curve in the space-time M, or for the limit on $\mathscr{I}$ of a tensor field on M. Such questions cannot be formulated directly for $\mathcal{S}$.

We conclude this section with a brief discussion of conserved quantities. Suppose we take the average, over the 2-sphere of the η, of any expression obtained from $\boldsymbol{\alpha}, \beta, \gamma$, and $\boldsymbol{\delta}$, e.g.,

$$T_{ab} = \mathrm{Av}\left[(\partial_{(a}\, \boldsymbol{\alpha})\partial_{b)}\boldsymbol{\gamma}\right]. \tag{88}$$

Equation (88) associates a tensor at Λ with any asymptotically flat initial-data set. Furthermore, the quantity T_{ab} is, in a sense, conserved, for it is certainly invariant under (generalized) translations. Of course, there exist a large number of quantities such as (88), few of which will be interesting physically.

In special relativity, a conserved quantity is normally a constant tensor field on Minkowski space. For example, the energy momentum of a closed system is a constant vector field. The Poincaré group, acting on Minkowski space, takes constant tensor fields to constant tensor fields; but, of course, the translations do not change such tensor fields. Hence, it is the quotient group, the Lorentz group, which acts effectively on constant tensor fields. In short, conserved quantities in special relativity form finite-dimensional representations of the Lorentz group.

We consider again an average such as (88). Using (84)–(87), we have

$$\dot{T}_{ab} = S^c T_{abc}, \tag{89}$$

where

$$\begin{aligned} T_{abc} = \mathrm{Av}[&\eta_c(\partial_{(a}\beta)\partial_{b)}\boldsymbol{\gamma} + \beta \mathbf{h}_{c(a}\partial_{b)}\boldsymbol{\gamma} \\ &- \beta\, \eta_c\, \eta_{(a}\partial_{b)}\boldsymbol{\gamma} \\ &+ \eta_c(\partial_{(a}\boldsymbol{\delta})\,\partial_{b)}\boldsymbol{\alpha} + \boldsymbol{\delta}\mathbf{h}_{c(a}\partial_{b)}\boldsymbol{\alpha} - \boldsymbol{\delta}\; \eta_c\; \eta_{(a}\partial_{b)}\boldsymbol{\alpha}]. \end{aligned} \tag{90}$$

Similarly, $T_{abc} = s^d T_{abcd}$, where T_{abcd} is still another average involving $\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}$, and $\boldsymbol{\delta}$. By continuing in this way, one obtains an infinite sequence of tensors at Λ, with the evolution of each tensor in the sequence expressed in terms of the next. That is to say, beginning with a single average such as (88), one obtains in general an infinite-dimensional representation of the Lorentz group.

These remarks motivate the following definition: By a *conserved quantity* we shall understand a finite collection of tensors at Λ, each obtained by taking an average of an expression involving $\boldsymbol{\alpha}, \beta, \gamma$, and δ, such that the evolution (under φ) of each tensor in this collection can be expressed in terms of the values of the tensors in the collection and the vector s_a where φ is defined.

We give one example of a conserved quantity. Set

$$M = -\tfrac{1}{3}\,\mathrm{Av}\boldsymbol{\alpha}, \quad P_a = -\tfrac{1}{3}\,\mathrm{Av}\,(\beta\eta_a). \tag{91}$$

Then, from (84)–(87), we have

$$\dot{M} = s^a P_a, \quad \dot{P}_a = M s_a \tag{92}$$

Physically, the conserved quantity (M, P_a) is the energy-momentum of the gravitating system (sources plus gravitational field). This quantity was discovered originally by Arnowitt, Deser, and Misner.[5] The "rest mass" of the system is the single conserved quantity $M^2 - P^a P_a$. One can think of (M, P_a) as a "free 4-vector." Then M can be interpreted as the "component of this 4-vector perpendicular to the initial surface S used in (91)," and P_a as the component parallel to S.

While it is not difficult to justify the physical interpretation of (M, P_a), there are a number of other conserved quantities at spatial infinity in general relativity whose interpretations are more obscure. These quantities will be discussed in detail in a subsequent paper.

6. SUMMARY AND CONCLUSION

Suppose one wishes to introduce quantities in general relativity which describe the structure of a gravitating system as a whole. In flat space, such quantities are associated with the action of the Poincaré group; but, unfortunately, this action is lost[29] in the presence of curvature. One is thus led to consider only metrics which approach a Minkowski metric at infinity, in order that the Poincaré group, or at least some similar group, will emerge asymptotically. Roughly speaking, the faster the space–time metric approaches a Minkowski metric, the closer the asymptotic symmetry group will be to the Poincaré group. On the other hand, the faster the metric approaches a Minkowski metric, the less the information that will be available from the asymptotic behavior of the gravitational field. Thus, in order to obtain a useful asymptotic description of gravitation, one must require that the metric become flat sufficiently slowly that useful information is not lost, but sufficiently quickly that the information which is available can be given a reasonable interpretation. The notion of asymptotic flatness represents a practical compromise between these two effects.

In order to describe asymptotic behavior at spatial infinity, we introduce a three-dimensional spacelike surface S which carries initial data for the space–time. Using a conformal completion of S by a point Λ at infinity, one can describe asymptotic behavior of the initial data in terms of local behavior at Λ. The asymptotic gravitational field on S is completely characterized by four scalars $\alpha(\eta)$, $\beta(\eta)$, $\gamma(\eta)$, and $\delta(\eta)$, which depend on directions (η) at Λ. The addition of a constant to δ, or of a term linear in η to γ, has no physical significance.

The conditions for asymptotic flatness on initial data have the property that, if they are satisfied for one initial-data set S in space–time, then they are satisfied for all initial-data sets with appropriate asymptotic behavior relative to S. Furthermore, the structure at spatial infinity is deterministic: A knowledge of the asymptotic field at one instant of time (i.e., for one initial-data set) completely and uniquely determines the field at all times. One would therefore expect there to exist a set of equations which govern the evolution of the asymptotic field α, β, γ, and δ. These equations are (84)–(87). Since there is no preferred choice of an initial-data set in the general space–time, the evolution equations represent an integral part of the structure at spatial infinity.

Since the four scalars α, β, γ, and δ completely characterize the asymptotic gravitational field, and since these scalars need be determined only once (i.e., for a single asymptotically flat initial-data set in the space–time), the problem of analyzing spatial infinity reduces to that of studying these scalars and their evolution equations (84)–(87). The original four-dimensional space–time, Einstein's equation, etc. play no further role.

Thus, in the asymptotic limit, the rather complicated Einstein equation reduces to the simple Equations (84)–(87).

APPENDIX A: THE INITIAL-VALUE EQUATIONS

Let M be a four-dimensional manifold with a smooth metric g_{ab} of signature $(-,+,+,+)$ which satisfies Einstein's equation:

$$R_{ab} - \tfrac{1}{2} R g_{ab} = T_{ab}. \tag{A1}$$

In this appendix, we shall derive the well-known[9] equations for the initial-value formulation of (A1). The present derivation is, in my opinion, both conceptually and computationally simpler than the standard ones.

Roughly speaking, one sets data for (A1) at some initial time. This data consists of a pair of tensor fields subject to certain constraint equations obtained from (A1). One then derives from (A1) the evolution equations, which give the change in the data from one instant of time to the next.

Let t be a smooth scalar field on M whose gradient is everywhere timelike. Then the one-parameter family of surfaces S_t, defined by $t = \text{const}$, are spacelike and three-dimensional. These surfaces represent the "times" in the paragraph above. Let ξ^a denote the unit normal to this family of surfaces, and let $\varphi \xi^a$ be the connecting vector field from each surface S_t to nearby surfaces. (I.e., $\varphi = [-(\nabla_a t)(\nabla^a t)]^{-1/2}$, $\xi_a = \varphi \nabla_a t$.)

We wish to discuss differential geometry within the surfaces S_t. It is convenient to proceed in such a way that all tensors and tensor operations on the S_t involve, formally, tensors and operations in M. (For analogous procedures in somewhat different cases, see Ref. 30.) A tensor field $T^{a\cdots c}{}_{b\cdots d}$ on M will be said to be *on the* S_t if

$$\begin{aligned} &\xi_a T^{a\cdots c}{}_{b\cdots d} = 0, \ldots, \xi_c T^{a\cdots c}{}_{b\cdots d} = 0, \\ &\xi^b T^{a\cdots c}{}_{b\cdots d} = 0, \ldots, \xi^d T^{a\cdots c}{}_{b\cdots d} = 0. \end{aligned} \tag{A2}$$

In particular,

$$h_{ab} = g_{ab} + \xi_a \xi_b \tag{A3}$$

is a tensor field on the S_t. Note that this h_{ab} is positive-definite (for vectors on the S_t), that the indices of a tensor on the S_t can be raised or lowered with either the four-dimensional metric g_{ab} or the metric h_{ab} with the same result, and that h_{ab} is a projection: $h_{am} h_b{}^m = h_{ab}$. Note also that any tensor field on M, by projecting parallel and perpendicular to ξ^a can be expressed completely in terms of tensors on the S_t. If $T^{a\cdots c}{}_{b\cdots d}$ is any tensor field on the S_t, then so is

$$D_e T^{a\cdots c}{}_{b\cdots d} = h_e^r h^a{}_m \cdots h^b{}_n h_c^p \cdots h_d^q \nabla_r T^{m\cdots n}{}_{p\cdots q}. \tag{A4}$$

The derivative operator D_a, defined by (A4), satisfies the usual rules (Liebnitz rule, absence of torsion, commutes with contraction). Furthermore, the derivative of the metric is zero:

$$D_a h_{bc} = 0. \tag{A5}$$

Hence, D_a is the covariant derivative for tensors on the S_t.

We define the *extrinsic curvature* of our family of surfaces by

$$p^{ab} = \nabla^a \xi^b + \varphi^{-1} \xi^a \nabla^b \varphi + \varphi^{-1} \xi^a \xi^b (\xi^m \nabla_m \varphi). \tag{A6}$$

Via the identity $\nabla_{[a} \xi_{b]} = -\varphi^{-1} \xi_{[a} \nabla_{b]} \varphi$ (which follows from the fact that $\varphi^{-1} \xi_b$ is a gradient), we see that p^{ab} is a symmetric tensor field on the S_t.

The positive-definite metric h_{ab} and the symmetric p^{ab} constitute the initial data for Einstein's equation (A1).

The first constraint equation is obtained by taking the divergence of $(p^{ab} - ph^{ab})$, where $p = p_m{}^m$:

$$D_b(p^{ab} - ph^{ab}) = h^{am} h^{bn} \nabla_b (p_{mn} - ph_{mn}). \tag{A7}$$

By substituting (A6) into the right side of (A7) and using (A1),

$$D_b(p^{ab} - ph^{ab}) = h^{am} \xi^n T_{mn}. \tag{A8}$$

The second constraint equation relates the curvature of the S_t to the extrinsic curvature. To evaluate the Riemann tensor $\mathcal{R}_{abcd}$ of the S_t, we commute D derivatives. Let k_c be an arbitrary vector field on the S_t. Then

$$\begin{aligned}
\tfrac{1}{2} \mathcal{R}_{abcd} k^d &= D_{[a} D_{b]} k_c = h_{[a}{}^m h_{b]}{}^n h_c{}^p \nabla_m \\
&\quad \times [h_n{}^r h_p{}^q \nabla_r k_q] \\
&= \tfrac{1}{2} h_a{}^m h_b{}^n h_c{}^p R_{mnpq} k^q \\
&\quad + h_{[a}{}^m h_{b]}{}^n h_c{}^p (\nabla_r k_q) \\
&\quad \times [h_p{}^q \nabla_m h_n{}^r + h_n{}^r \nabla_m h_p{}^q] \\
&= \tfrac{1}{2} h_a{}^m h_b{}^n h_c{}^p h_d{}^q R_{mnpq} k^d \\
&\quad - p_{c[a} p_{b]d} k^d,
\end{aligned} \tag{A9}$$

where we have used $\xi^m k_m = 0$. Since k_c is arbitrary, we have, substituting (A6),

$$\mathcal{R}_{abcd} = h_a{}^m h_b{}^n h_c{}^p h_d{}^q R_{mnpq} - 2 p_{c[a} p_{b]d}. \tag{A10}$$

By contracting (A10) twice and using (A1), we obtain the second constraint equation[16]:

$$\mathcal{R} - p^{mn} p_{mn} + p^2 = 0. \tag{A11}$$

Finally, we obtain the evolution equations. A dot, affixed to a tensor field on the S_t, will denote the Lie derivative of that field by $\varphi \xi^a$. Thus, the dot denotes the rate of change from each surface S_t to the next. We have

$$\dot{h}_{ab} = \pounds_{\varphi\xi} h_{ab} = \varphi \xi^m \nabla_m (h_{ab}) + h_{am} \nabla_b (\varphi \xi^m) + h_{bm} \nabla_a (\varphi \xi^m). \tag{A12}$$

By substituting (A3) and using (A6),

$$\dot{h}_{ab} = 2\varphi p_{ab}. \tag{A13}$$

Similarly,

$$\dot{p}^{ab} = \varphi \xi^m \nabla_m p^{ab} - p^{am} \nabla_m (\varphi \xi^b) - p^{bm} \nabla_m (\varphi \xi^a). \tag{A14}$$

By substituting (A6) and using (A11) and (A1),

$$\dot{p}^{ab} = D^a D^b \varphi - 2\varphi p^a{}_m p^{bm} - \varphi p p^{ab} - \varphi \mathcal{R}^{ab}. \tag{A15}$$

The initial data h_{ab} and p^{ab}, subject to the constraint equations (A7) and (A11), evolve according to (A13) and (A15).

APPENDIX B: A COMPUTATIONAL TECHNIQUE

The situation arises occasionally in general relativity that one wishes to determine the change in some quantity, defined originally in terms of the metric and perhaps other fields on a manifold, under a certain specified change in the metric. There is a simple, coordinate-independent method for calculating such effects. In this appendix, we illustrate this technique with two examples. The results are required at a number of points in Secs. 2 and 3.

Suppose we perform a conformal transformation

$$\tilde{g}_{ab} = \Omega^2 g_{ab}, \tag{B1}$$

where Ω is a positive scalar field on the n-dimensional manifold M. We ask for the effect of (B1) on, say, the Riemann tensor. Let ∇_a be the derivative operator with respect to g_{ab} and $\tilde{\nabla}_a$ with respect to $\tilde{g}_{ab}$. Then, since both derivative operators are torsion-free and satisfy the Liebnitz rule, there exists a tensor field $C^m{}_{ab} = C^m{}_{(ab)}$ on M such that

$$\begin{aligned}
\tilde{\nabla}_m T^{a\cdots c}{}_{b\cdots d} &= \nabla_m T^{a\cdots c}{}_{b\cdots d} + C^a_{mn} T^{n\cdots c}{}_{b\cdots d} \\
&\quad + \cdots + C^c_{mn} T^{a\cdots n}{}_{b\cdots d} - C^n_{mb} T^{a\cdots c}{}_{n\cdots d} - \cdots \\
&\quad - C^n_{md} T^{a\cdots c}{}_{b\cdots n},
\end{aligned} \tag{B2}$$

for any tensor field $T^{a\cdots c}{}_{b\cdots d}$ on M. By substituting (B1) and (B2) into $\tilde{\nabla}_a \tilde{g}_{bc} = 0$, we obtain

$$C^m_{ab} = + \Omega^{-1} (\delta^m_a \nabla_b \Omega + \delta^m_b \nabla_a \Omega - g^{mn} g_{ab} \nabla_n \Omega). \tag{B3}$$

The equation relating the Riemann tensors $\mathbf{R}_{abcd}$ and $\tilde{R}_{abcd}$ of g_{ab} and $\tilde{g}_{ab}$, respectively, is obtained as follows. Fix a vector field k_c on M. Then

$$\tfrac{1}{2} \tilde{R}_{abc}{}^m k_m = \tilde{\nabla}_{[a} \tilde{\nabla}_{b]} k_c. \tag{B4}$$

By substituting (B2),

$$\begin{aligned}
\tfrac{1}{2} \tilde{R}_{abc}{}^m k_m &= \nabla_{[a} (\nabla_{b]} k_c - C^m_{b]c} k_m) \\
&\quad - C^n_{[ab]} (\nabla_n k_c - C^m_{nc} k_m) \\
&\quad - C^n_{c[a} (\nabla_{b]} k_n - C^m_{b]n} k_m) \\
&= \tfrac{1}{2} R_{abc}{}^m k_m - k_m \nabla_{[a} C^m_{b]c} + k_m C^n_{c[a} C^m_{b]n}.
\end{aligned} \tag{B5}$$

But k_c is arbitrary, and so

$$\begin{aligned}
\tilde{R}_{abc}{}^d &= R_{abc}{}^d - 2\nabla_{[a} C^m_{b]c} + 2 C^n_{c[a} C^m_{b]n} \\
&= R_{abc}{}^d + 2\Omega^{-1} g_{c[b} \nabla_{a]} \nabla^d \Omega \\
&\quad - 2\Omega^{-1} \delta^d{}_{[b} \nabla_{a]} \nabla_c \Omega
\end{aligned}$$

$$+ 4\Omega^{-2} (\nabla_c\Omega)\, \delta^d{}_{[b}\nabla_{a]}\Omega - 4\Omega^{-2} \times (\nabla^d\Omega) g_{c[b}\nabla_{a]}\Omega - 2\Omega^{-2} g_{c[a}\delta_{b]}{}^d \times (\nabla^m\Omega)(\nabla_m\Omega) \tag{B6}$$

where, in the second step, we have used (B3). Similarly, if is a scalar field on M of dimensions[12] sec, so $\tilde{\varphi} = \Omega\varphi$, then

$$\begin{aligned}\tilde{\Box}\tilde{\varphi} &= \tilde{g}^{ab}\tilde{\nabla}_a\tilde{\nabla}_b(\Omega\varphi) = \Omega^{-2}g^{ab}\tilde{\nabla}_a\nabla_b(\Omega\varphi)\\ &= \Omega^{-2}g^{ab}[\nabla_a\nabla_b(\Omega\varphi) - C^m_{ab}\nabla_m(\Omega\varphi)]\\ &= \Omega^{-1}\Box\varphi + \Omega^{-2}\varphi\Box\Omega + n\Omega^{-2}(\nabla_m\Omega)(\nabla^m\varphi)\\ &\quad + (n-2)\,\Omega^{-3}\varphi(\nabla^m\Omega)(\nabla_m\Omega).\end{aligned} \tag{B7}$$

As a second example of essentially the same technique, we derive a perturbation equation. Let $g_{ab}(t)$ be a one-parameter family of non singular metrics on the n-dimensional manifold M. Then, for each t, we may compute the Riemann tensor $R_{abcd}(t)$. We wish to evaluate, for example, $(d/dt)R_{abcd}(t) = \dot{R}_{abcd}$. Since the derivative operator $\nabla_a(t)$ is torsion-free and satisfies the Liebnitz rule for each t, there is a tensor field $C^m{}_{ab}(t) = C^m{}_{(ab)}(t)$ on M, such that

$$\begin{aligned}(\nabla_m T^{a\cdots c}{}_{b\cdots d})^{\cdot} &= \nabla_m \dot{T}^{a\cdots c}{}_{b\cdots d} + C^a_{mn} T^{n\cdots c}{}_{b\cdots d}\\ &\quad + \cdots + C^c_{mn} T^{a\cdots n}{}_{b\cdots d} - C^n_{mb} T^{a\cdots c}{}_{n\cdots d} - \cdots\\ &\quad - C^n_{md} T^{a\cdots c}{}_{b\cdots n}\end{aligned} \tag{B8}$$

for any tensor field $T^{a\cdots c}{}_{b\cdots d}(t)$ on M. Taking d/dt of the equation $\nabla_a(t)g_{bc}(t) = 0$, we have, using (B8),

$$C^m_{ab} = -\tfrac{1}{2}g^{mn}(\nabla_a\dot{g}_{bn} + \nabla_b\dot{g}_{an} - \nabla_n\dot{g}_{ab}). \tag{B9}$$

Thus, for an arbitrary, t-independent vector field k_c, we have

$$\begin{aligned}\tfrac{1}{2}\dot{R}_{abc}{}^m k_m &= (\nabla_{[a}\nabla_{b]}k_c)^{\cdot}\\ &= \nabla_{[a}(C^m_{b]c}k_m) + C^m_{[ab]}\nabla_m k_c + C^m_{c[a}\nabla_{b]}k_c\\ &= k_m\nabla_{[a}C^m_{b]c}.\end{aligned} \tag{B10}$$

By substituting (B9) and using the fact that k_c is arbitrary, we obtain

$$\dot{\mathcal{R}}_{abc}{}^d = \tfrac{1}{2}g^{dm}[-2\nabla_{[a}\nabla_{b]}\dot{g}_{cm} - \nabla_a\nabla_c\dot{g}_{bm} + \nabla_b\nabla_c\dot{g}_{am} + \nabla_a\nabla_m\dot{g}_{bc} - \nabla_b\nabla_m\dot{g}_{ac}]. \tag{B11}$$

Thus, for example,

$$\dot{\mathcal{R}}_{ab} = \nabla^m\nabla_{(a}\dot{g}_{b)m} - \tfrac{1}{2}\nabla_a\nabla_b(g^{mn}\dot{g}_{mn}) - \tfrac{1}{2}\nabla^m\nabla_m\dot{g}_{ab}. \tag{B12}$$

---

1 H. Bondi, M. G. J. Van Der Berg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).

2 R. Sachs, Proc. Roy. Soc. (London) **A270**, 193 (1962).

3 See, for example, F. A. E. Pirani, article in *Gravitation*, edited by L. Witten (Wiley, New York, 1962), p. 199.

4 R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

5 R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **118**, 1100 (1960); **121**, 1556 (1961); **122**, 997 (1961).

6 No relation is known relating the energy-momentum as determined at null infinity to that at spatial infinity. The most promising conjecture is that the energy-momentum at future null infinity, in the limit to the past, is equal, in some sense, to the (conserved) energy-momentum at spatial infinity.

7 R. Geroch, J. Math. Phys. **11**, 2580 (1970).

8 It is, of course, permissible to have sources everywhere, provided that the stress-energy goes to zero asymptotically at an appropriate rate.

9 A. Lichnerowicz, *Theories Relativistes de la Gravitation et de L'electromagnetisme* (Masson, Paris, 1955).

10 See, for example, R. Geroch, J. Math. Phys. **11**, 437 (1970).

11 Of course, all quantities at spatial infinity are invariant under translations. The conserved quantities will be defined in terms of their behavior under the Lorentz group.

12 By the *dimension* of a tensor field, we mean sec^n, where n is the following number: [power of Ω by which that tensor is to be multiplied under (10)] − (number of covariant indices) + (number of contravariant indices). Thus, the metric itself is dimensionless, and the dimension is unchanged by the raising and lowering of indices and by contraction. This formula gives correctly the usual, physical dimensions of quantities in general relativity.

13 By contracting (25), we have $\partial^2\varphi + 2\varphi = 0$. More generally, if $\mathbf{f}$ is of the form $\mathbf{f} = \mathbf{Q}_{a_1\ldots a_n}\eta^{a_1}\cdots\eta^{a}{}_n$, where $\mathbf{Q}_{a_1\ldots a_n}$ is trace free, then $\partial^2\mathbf{f} + n(n+1)\mathbf{f} = 0$. Of course, this formula is well known in the study of spherical harmonics.

14 Since a system can have no net magnetic charge, why do we not require in addition that $\mathbf{B}_a\eta^a = 0$? Because this condition is not preserved by (26).

15 A simple geometrical interpretation of the extrinsic curvature is as follows. Let γ be a geodesic within the surface S, and let μ^a be the unit tangent vector to this curve. Then γ, considered as a curve in the space–time M, will not be a geodesic. In fact, we have $\mu^m\nabla_m\mu^a = \xi^a(\mu^m\mu^n p_{mn})$, where ξ^a is the unit normal to S.

16 Our conventions are as follows: $\nabla_{[a}\nabla_{b]}k_c = \frac{1}{2}R_{abc}{}^d k_d$, $R_{ab} = R_{amb}{}^m$.

17 Note that limits of tensor fields at the point Λ are well defined in this case because the manifold $\bar{S}$ has a C^1 differentiable structure.

18 Since $\tilde{h}_{ab}$ is continuous at Λ, $\mathbf{h}_{ab}$ is independent of η.

19 Equation (52) is the usual relationship between the Ricci and Riemann tensors in three dimensions.

20 That this difficulty is actually caused by the presence of curvature can be seen from the right side of (55).

21 More precisely, this expression is to represent the conformal factor to first order in displacements from Λ.

22 Why not merely demand that ω be linear in η? In order to obtain a conformal completion of S, it was necessary to make a choice of a conformal factor Ω. The freedom in ω must reflect the freedom originally available in Ω.

23 In fact, the two groups are isomorphic. However, since the groups operate in different asymptotic limits, this fact does not appear to be very significant.

24 This assertion is easy to prove. If $\frac{1}{4}\mathbf{D} = \mathbf{Q} + \mathbf{Q}_{ab}\eta^a\eta^b + \mathbf{Q}_{abc}\eta^a\eta^b\eta^c + \cdots$, then $\gamma = \frac{1}{2}\mathbf{Q} - \frac{1}{4}\mathbf{Q}_{ab}\eta^a\eta^b + \cdots + (-n^2 - n + 2)^{-1}\mathbf{Q}_{a_1\ldots a_n}\eta^{a_1}\cdots\eta^{a_n} + \cdots$. See Footnote 13.

25 The direct proof of this statement is rather inelegant. Expand $\mathbf{E}'_{ab}$ as in footnote 24.

26 The quantities $\mathbf{E}_a$ and $\mathbf{B}_a$ can be interpreted, more abstractly, as vector fields on the tangent bundle of S.

27 That is to say, the group of isometries on a complete, simply-connected, three-dimensional manifold with a positive-definite metric of constant curvature is the Lorentz group (plus a reflection).

28 One must verify that this procedure is consistent, i.e., that, if one transports α, β, γ, and δ around a closed curve, via (84)–(87), then these quantities return to the starting point with their original values. This verification is not difficult.

29 The Poincarè group either reduces to the group with only an identity element, or else becomes infinite-dimensional, depending on one's approach. If one asks for isometries in curved space, then, in general, there is only one: the identity. If one asks for an "invariance group," he may obtain the group of all diffeomorphisms on the manifold.

30 R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).

Scattering Theory in Fock Space*

Eduard Prugovečki
Department of Mathematics, University of Toronto, Toronto 181, Canada
(Received 3 December 1971; Revised Manuscript Received 29 February 1972)

It is shown that the wave operators which can be introduced in quantum field theories with cutoff interactions can be represented by wave transformers acting on states represented by statistical operators. Integral representations can be found for these wave transformers, which then form the basis of a time-independent scattering theory for such quantum field theories. It is pointed out that this leads to a time-independent perturbation theory in Fock space.

1. INTRODUCTION

The time-dependent scattering theory for the nonrelativistic two-body problem in Hilbert space $\mathcal{H}$ is based on the concept of "strong" wave operators $\Omega_\pm$ associated with the pair H_0 and H of self-adjoint operators in $\mathcal{H}$, representing the free and interacting Hamiltonians, respectively. In the most common case of potential scattering H_0 has only an absolutely continuous spectrum for large classes of potentials. Then one defines

$$\Omega_\pm = \underset{t\to\mp\infty}{\text{s-lim}}\, W(t), \tag{1.1}$$

where

$$W(t) = V^*(t)U(t), \qquad U(t) = \exp(-iH_0t), \quad V(t) = \exp(-iHt), \tag{1.2}$$

and one can prove existence of the above strong limit for many[1] short-range potentials. A time-independent approach can be based on integral representations[1]

$$W_{\pm\epsilon} = \int_{-\infty}^{+\infty} \frac{\mp i\epsilon}{H-\lambda\mp i\epsilon}\, dE_\lambda, \qquad \epsilon > 0, \tag{1.3}$$

where E_λ is the spectral function of H_0. These integrals involve a concept of Riemann–Stieltjes integration of operator-valued functions with respect to spectral measures, which has been developed in several different contexts,[2-6] and which is quite natural and elementary in its most basic form.[1] It can be shown[6] that integrals of the type occuring in (1.3) exist under very general circumstances, irrespective of any possible time-dependent definition[6-8] of such integrals. However, if (1.1) holds then it has been derived independently by several authors[7-9] that

$$\Omega_\pm = \underset{\epsilon\to+0}{\text{s-lim}}\, W_{\pm\epsilon}, \tag{1.4}$$

and that most of the usual formulas for the S matrix, T matrix, and wave operators encountered in physical literature on time-independent two-body scattering theory can be provided with counterparts in Hilbert space, which are derivable from (1.4).[1,9-11]

General approaches to time-dependent scattering theory which would be applicable also to long range potentials have been recently proposed by several authors.[12-16] They disagree in their basic physical assumptions, but in case of potential scattering for long-range forces they all lead to a modification of the definition (1.1) of $\Omega_\pm$ which consists in replacing $U(t)$ in (1.2) by

$$U'(t) = \exp[-iH_0t + G_t(H_0)], \tag{1.5}$$

where $G_t(H_0)$ is a suitably chosen function of H_0 and t.[12,16] A corresponding modification of the relation (1.3) for $W_{\pm\epsilon}$ can be then derived for this more general case.[17]

It has been recently observed that a method first applied[18] when defining asymptotic creation and annihilation operators for quantum field theories with interactions which are regular perturbations of the free Hamiltonian, can be also applied to many interactions which are integrals with space cutoff of Wick ordered polynomials in the fields[19-21] with momentum cutoffs (except in the: $\varphi^4{:}_2$ case, when the ultraviolet cutoff is not needed). This leads to natural time-dependent definitions (cf. Sec. 2) of wave operators $W_\pm$ for such models. Furthermore, it can be shown[22] that such wave operators have the expected physical behaviour in relation to spin and momentum measurements with which scattering experiments are usually concerned.

In Sec. 2 of this paper we formulate two essential features of these scattering theories in Fock space in the form of two simple postulates. Then we show that other relevant features of such theories (such as intertwining relations, spectrum of H, etc.) can be derived exclusively from these two postulates.

In Sec. 3 we develop the time-independent counterpart of this scattering theory by introducing the concept of wave transformers $\underline{W}_\pm$ on the Hilbert-Schmidt class $\mathcal{B}_2(\mathcal{F})$ in the C^*-algebra of bounded operators on Fock space $\mathcal{F}$. We derive for $\underline{W}_\pm$ relations analogous to (1.3) and (1.4). The general theorems on Abelian additive one-parameter group of unitary operators in $\mathcal{B}_2(\mathcal{F})$ required in these derivations are stated and proven in the Appendix.

In Sec. 4 we indicate what is the significance of the derived results for perturbation theory.

2. GENERAL FEATURES OF TIME-DEPENDENT SCATTERING THEORY IN FOCK SPACE

Let us denote by $\mathcal{F}$ the Fock space of a finite number of massive free relativistic or nonrelativistic fields described by the creation and annihilation sesquilinear forms $a^\#(\mathbf{k},\sigma,\nu)$; here the index $\nu = 1,\ldots,\nu_0$ distinguishes between the different kinds of considered particles and antiparticles (a total of ν_0 kinds), while $\mathbf{k}\in\mathbb{R}^3$ and $\sigma\in\mathbb{R}^1$ are the 3-momentum and spin variables, respectively; $a^\#$ stands for a and a^*. A very economical notation is obtained if we introduce the closed operators

$$a^\#(f) \supseteq \int_{\mathbb{R}^5} a^\#(\mathbf{k},\sigma,\nu) f(\mathbf{k},\sigma,\nu)\, d\mu_0, \qquad f\in\mathcal{F}^{(1)} \equiv L^2_{\mu_0}(\mathbb{R}^5), \tag{2.1}$$

where μ_0 is an adequately chosen σ-finite measure on the Borel sets of $\mathbb{R}^5$ (cf. Refs. 21, 22) the exact form of which is, however, inessential to our considerations. It suffices to say that $\mathcal{F}^{(1)}$ can be considered to be the direct sum of all ν_0 one-particle (antiparticle) Hilbert spaces.

We shall denote by Φ_0, $\|\Phi_0\|^2 = \langle\Phi_0|\Phi_0\rangle = 1$, the Fock vacuum, where $\langle\cdot|\cdot\rangle$ is the inner produce in $\mathcal{F}$. We take the free Hamiltonian H_0 to be the usual generator of time translations in Fock space, so that $H_0\Phi_0 = 0$. We do not have, however, to specify the full Hamiltonian H beyond requiring that it satisfy certain conditions, which we formulate below in the form of two postulates. It has to be pointed out that these postulates are satisfied by all the Hamiltonians with cutoff interactions considered thus far[19–21] in the present time-dependent scattering theory framework for constructive quantum field theory.

Postulate A: The Hamiltonian H is self-adjoint and has an eigenvector Φ (with eigenvalue η) such that $\langle \Phi | \Phi_0\rangle \neq 0$.

Postulate B: There are closed operators $a_\pm^\#(f)$ for each $f \in \mathfrak{F}^{(1)} \equiv L^2_{\mu_0}(\mathbb{R}^5)$ such that for any vector Ψ in the domain $\mathfrak{D}_H$ of H

$$a_\pm^\#(f)\,\Psi = \underset{t\to\mp\infty}{\text{s-lim}}\, W(t) a^\#(f) W^*(t)\Psi, \qquad (2.2)$$

where $W(t)$ is defined in (1.2) in terms of H_0 and H.

As an immediate and straightforward consequence of (2.2) and the fact that $a^\#(f)$ are free fields, we obtain the following result (cf. Refs. 18, 21, 24).

Lemma 2.1: The asymptotic fields $a_\pm^\#(f)$ satisfy the same canonical commutation and anticommutation relations as their free field counterparts $a^\#(f)$ do, and $a_\pm(f)\Phi = \mathbf{0}$ for all $f \in \mathfrak{F}^{(1)}$.

Let us denote by $\mathfrak{F}_+$ and $\mathfrak{F}_-$ the closed subspaces of $\mathfrak{F}$ which are spanned by all possible polynomials in a_+^* and a_-^*, respectively, applied to Φ. Then in view of Lemma 2.1 we can make the following statement (cf. Refs. 21, 24).

Theorem 2.1: There are partial isometries $W_\pm$ with initial domain $\mathfrak{F}$ and final domains $\mathfrak{F}_\pm$ such that $W_\pm \Phi_0 = \Phi$ and

$$W_\pm a^*(f_1)\cdots a^*(f_n)\Phi_0 = a_\pm^*(f_1)\cdots a_\pm^*(f_n)\Phi \qquad (2.3)$$

for any $f_1, \cdots, f_n \in \mathfrak{F}^{(1)}$ and $n = 1, 2, \cdots$.

The statement that $W_\pm$ are partial isometries, made in the above theorem, holds if Φ is normalized in $\mathfrak{F}$ and if we choose in $\mathfrak{F}_\pm$ the same inner product as in $\mathfrak{F}$. However, we shall see later that

$$\underset{t\to+\infty}{\text{w-lim}}\, W(t)|\Phi_0\rangle\langle\Phi_0| W^*(t)P_\pm = e^{-\Lambda}|\Phi\rangle\langle\Phi|, \qquad (2.4)$$

where $P_\pm$ are the projectors onto $\mathfrak{F}_\pm$, respectively, and Λ plays the role of a wave renormalization constant[22]:

$$\Lambda = -\ln|\langle\Phi|\Phi_0\rangle|^2. \qquad (2.5)$$

Hence, it turns out to be more convenient to introduce in $\mathfrak{F}_\pm$ a new inner product (cf. also Lemma 3.2)

$$\langle\cdot|\cdot\rangle_\pm = e^{+\Lambda}\langle\cdot|\cdot\rangle \qquad (2.6)$$

and have $\Phi \in \mathfrak{F}_\pm$ normalized with respect to this new inner product, thus retaining $W_\pm$ as unitary operators from $\mathfrak{F}$ to $\mathfrak{F}_\pm$:

$$\|\Phi\|^2 = e^{-\Lambda}\|\Phi\|_\pm^2 = e^{-\Lambda}. \qquad (2.7)$$

Lemma 2.2: The wave operators $W_\pm$ intertwine $a^\#(f)$ and $a_\pm^\#(f)$, i.e.,

$$W_\pm a^\#(f)\Psi = a_\pm^\#(f) W_\pm \Psi \qquad (2.8)$$

for any $\Psi \in W_\pm^*\mathfrak{D}_H$, as well as H_0 and H-η, i.e.,

$$\exp[i(H-\eta)s]W_\pm = W_\pm \exp(iH_0 s) \qquad (2.9)$$

for all $s \in \mathbb{R}^1$.

Proof: The relation (2.8) is obtained for creators $a^*(f)$ and for $\Psi = \Phi_0$ as an immediate consequence of (2.3) and the fact that $W_\pm^* W_\pm = \mathbf{1}$ since $W_\pm$ are partial isometries with initial domain $\mathfrak{F}$. The generalization to $\Psi = a^*(f_1)\cdots a^*(f_n)\Phi_0$ is then derived by induction in n based on the same method. By taking closures we extend the result to all $\Psi \in W_\pm^*\mathfrak{D}_H$ and, since $a(f) = [a^*(\bar f)]^*$, also to all annihilators $a(f)$.

Using Postulate B and the fact that

$$\exp(iH_0 s)a^\#(f)\exp(-iH_0 s) = a^\#(\exp(\pm i\omega s)f), \qquad (2.10)$$

where

$$\omega(\mathbf{k}, \nu) = [\mathbf{k}^2 + m_\nu^2]^{1/2} \qquad (2.11)$$

and m_ν is the rest mass of particles of kind ν, we can easily justify each step in the following calculation, in which $\Psi \in \mathfrak{D}_H \cap \mathfrak{F}_\pm$:

$$\begin{aligned}
&\exp(iHs)a_\pm^\#(f)\exp(-iHs)\Psi \\
&= \exp(iHs)W_\pm a^\#(f)W_\pm^*\exp(-iHs)\Psi \\
&= \underset{t\to\mp\infty}{\text{s-lim}}\exp[iH(s+t)]\exp(-iH_0 t)a^\#(f) \\
&\quad\times \exp(iH_0 t)\exp[-iH(s+t)]\Psi \\
&= \underset{\tau\to\mp\infty}{\text{s-lim}}\exp(iH\tau)\exp(-iH_0\tau)a^\#(f) \\
&\quad\times \exp(\pm i\omega s)f)\exp(iH_0\tau)\exp(-iH\tau)\Psi \\
&= a_\pm^\#(\exp(\pm i\omega s)f)\Psi = W_\pm a^\#(\exp(\pm i\omega s)f)W_\pm^*\Psi \\
&= W_\pm \exp(iH_0 s)a^\#(f)\exp(-iH_0 s)W_\pm^*\Psi.
\end{aligned}$$

From the above relation we easily derive by algebraic manipulations the fact that

$$[a^\#(f), W_\pm^*\exp(-iHs)W_\pm\exp(iH_0 s)]_-\Psi' = 0 \qquad (2.12)$$

for the dense set of vectors Ψ' from $W_\pm^*\mathfrak{D}_H$. Since the family $\{a^\#\}$ of all smeared creators and annihilators forms an irreducible set of operators (cf. Ref. 23, Theorem 4-5), (2.12) implies that

$$W_\pm^*\exp(-iHs)W_\pm\exp(iH_0 s) = \gamma(s)\mathbf{1},$$

where $\gamma(s)$ is a complex function with $|\gamma(s)| \equiv 1$. By applying the above relation to Φ_0, we obtain $\gamma(s) = \exp(-i\eta s)$, i.e., (2.9) holds.

The importance of the proof of the above lemma to our later considerations lies in its having the feature of relying completely on Postulates A and B, rather than on the study[18,21,24] of the behavior in t of the commutator

$$[H_I, a^*(\exp(i\omega t)f)]_-, \qquad H_I = H - H_0, \qquad (2.13)$$

for which a detailed knowledge of the interaction energy density $H_I(\mathbf{x})$ in terms of the fields is necessary. Due to this fact we can immediately infer the following results without having to specify the field-theoretical structure of H_I.

Theorem 2.2: The Hamiltonian H leaves $\mathfrak{F}_\pm$ invariant. Furthermore, $H - \eta 1$ has the same spectrum in $\mathfrak{F}_\pm \ominus \mathfrak{F}_\pm^{(0)}$ as H_0 has in $\mathfrak{F} \ominus \mathfrak{F}^{(0)}$, and (2.4) holds.

In the above theorem, η is the eigenvalue of H for the eigenvector Φ, while $\mathfrak{F}_\pm^{(0)}$ and $\mathfrak{F}^{(0)}$ denote the one-dimensional spaces spanned by Φ and Φ_0, respectively.

The first part of Theorem 2.2 is a direct consequence of (2.9). Moreover, since $W_\pm \Phi_0 = \Phi$, we get that $W_\pm$ maps $\mathfrak{F} \ominus \mathfrak{F}_\pm^{(0)}$ onto $\mathfrak{F}_\pm \ominus \mathfrak{F}_\pm^{(0)}$. Since $(H - \eta)\Phi = \mathbf{0}$ and H is self-adjoint, $H - \eta 1$ leaves $\mathfrak{F}_\pm \ominus \mathfrak{F}_\pm^{(0)}$ invariant, and by (2.9) it must have on $\mathfrak{F}_\pm \ominus \mathfrak{F}_\pm^{(0)}$ the same spectrum as H_0 has on $\mathfrak{F} \ominus \mathfrak{F}^{(0)}$. The fact that this spectrum is absolutely continuous can be then used to show that (2.4) holds. This can be done in the same manner as in Lemma 3.1 of Ref. 22.

3. INTEGRAL REPRESENTATIONS FOR WAVE TRANSFORMERS

It is well known[20,21] that if the interaction density $H_I(\mathbf{x})$ determining H_I contains pure creation terms, then the strong limit (1.1) does not exist despite the existence of the limit in (2.2). This is essentially due to the fact that in such cases the limits

$$\operatorname*{s-lim}_{t\to\mp\infty} W(t)\Phi_0 = \operatorname*{s-lim}_{t\to\mp\infty} \exp(iHt)\Phi_0 \tag{3.1}$$

do not exist since Φ_0 has a nonzero component in the absolutely continuous subspace of H. It is then natural to investigate the behavior of the operators $W_{\pm\epsilon}$ defined in (1.3) (which, we recall, exist whenever H and H_0 are self-adjoint) in the limit $\epsilon \to +0$, rather than the strong limit of $W(t)$ for $t \to \mp\infty$. Unfortunately, the limits of $W_{\pm\epsilon} a^\# W_{\pm\epsilon}^*$ cannot be directly related to the limits in (2.2) on account of the fact that the same value of t appears on both sides of $a^\#$ in (2.2). However, this difficulty can be by-passed if we consider the expression $W(t)a^\# W^*(t)$ to be a *transformation* of $a^\#$ rather than just the multiplication of $a^\#$ by $W(t)$ and $W^*(t)$ from left and right, respectively.

Let us denote by $\mathfrak{F}^{(n)}$ the closed subspace of $\mathfrak{F}$ which is spanned by all n-particle states,

$$\mathfrak{F}^{(n)} = [\{a^*(h_1)\cdots a^*(h_n)\Phi_0 | h_1, \ldots, h_n \in \mathfrak{F}^{(1)}\}], \tag{3.2}$$

for $n = 1, 2, \cdots$. We define in a similar manner

$$\mathfrak{F}_\pm^{(n)} = [\{a_\pm^*(h_1)\cdots a_\pm^*(h_n)\Phi | h_1, \ldots, h_n \in \mathfrak{F}^{(1)}\}]. \tag{3.3}$$

We obviously have

$$\mathfrak{F} = \bigoplus_{n=0}^{\infty} \mathfrak{F}^{(n)}, \quad \mathfrak{F}_\pm = \bigoplus_{n=0}^{\infty} \mathfrak{F}_\pm^{(n)}. \tag{3.4}$$

If $P^{(n)}$ and $P_\pm^{(n)}$ denote the projectors on $\mathfrak{F}^{(n)}$ and $\mathfrak{F}_\pm^{(n)}$ respectively, note that $a^\#(f)P^{(n)}$ are bounded operators for all $n = 0, 1, 2, \cdots$ and $f \in \mathfrak{F}^{(1)}$.

Lemma 3.1: If A is an element of the class $\mathfrak{B}_2(\mathfrak{F})$ of all Hilbert-Schmidt operators in $\mathfrak{F}$, so that

$$\sum_{n=0}^{\infty} \|P^{(n)} A\|_2^2 < \infty, \tag{3.5}$$

then $a^\#(f)P^{(m)}A$ and $(AP^{(m)}a^\#(f))^{**}$ also belong to $\mathfrak{B}_2(\mathfrak{F})$ for any $f \in \mathfrak{F}^{(1)}$.

Proof: We can obviously write

$$A = \sum_{n=0}^{\infty} P^{(n)}A = \sum_{n=0}^{\infty} AP^{(n)}. \tag{3.6}$$

From (3.6) and the fact that $\mathfrak{B}_2(\mathfrak{F})$ is a two-sided ideal in $\mathfrak{B}(\mathfrak{F})$ (cf. Appendix) and that there is a constant C_f for which

$$\|a^\#(f)P^{(m)}\| \le (m+1)^{1/2}C_f, \tag{3.7}$$

we easily infer the desired result.

Lemma 3.2: If $A \in \mathfrak{B}_2(\mathfrak{F})$, then

$$\operatorname*{w-lim}_{t\to\pm\infty} P_\pm W(t)P^{(m)}AP^{(n)}W^*(t)P_\pm = W_\pm P^{(m)}AP^{(n)}W_\pm^*, \tag{3.8}$$

for any $m, n = 0, 1, 2, \cdots$.

Proof: We shall prove (3.8) by induction in m and n.

The case $m = n = 0$ has been already established, since (3.8) for $m = n = 0$ is equivalent to relation (2.4), which was proven in Theorem 2.2.

Let us assume that (3.8) is true for some integers m and n. Then for any $f \in \mathfrak{F}^{(1)}$ we get

$$\operatorname*{w-lim}_{t\to\mp\infty} P_\pm W(t)a^*(f)P^{(m)}AP^{(n)}W^*(t)P_\pm = W_\pm a^*(f)P^{(m)}AP^{(n)}W_\pm \tag{3.9}$$

by using (3.8) and the relations (2.8), written in the form

$$\operatorname*{s-lim}_{t\to\mp\infty} P_\pm W(t)a^\#(h)W^*(t)P_\pm\Psi = W_\pm a^\#(h)W_\pm^*\Psi, \quad \Psi \in \mathfrak{D}_H, \tag{3.10}$$

to establish (3.8) for matrix elements of the given expressions taken for vectors $\Psi_1, \Psi_2 \in \mathfrak{D}_H$, and then extending to arbitrary $\Psi_1, \Psi_2 \in \mathfrak{F}$ by virtue of the fact that $a^*(f)P^{(m)}AP^{(n)}$ is bounded (cf. Lemma 3.1).

Now, any $\Psi^{(m+1)} \in \mathfrak{F}^{(m+1)}$ can be approximated arbitrarily well in norm by a finite linear combination of vectors of form $a^*(h)\Psi^{(m)}$ with $\Psi^{(m)} \in \mathfrak{F}^{(m)}$. Hence $|\Phi^{(m+1)}\rangle\langle\Psi^{(n)}|$ can be approximated arbitrarily well in operator norm by finite linear combinations of operators of the form

$$a^*(f)|\Phi^{(m)}\rangle\langle\Psi^{(n)}|. \tag{3.11}$$

Since $\|W(t)\| = \|W_\pm\|_\pm = 1$, we can use the aforementioned fact to extend (3.8) to operators $P^{(m+1)}AP^{(n)}$ of form (3.11).

Any $P^{(m+1)}AP^{(n)} \in \mathfrak{B}_2(\mathfrak{F})$ has a canonical decomposition[25]

$$P^{(m+1)}AP^{(n)} = \sum_{k=1}^{\infty} |\Phi_k^{(m+1)}\rangle \lambda_k \langle \Psi_k^{(n)}|, \qquad \sum_{k=1}^{\infty} \lambda_k^2 < \infty, \quad \lambda_k \geqslant 0, \tag{3.12}$$

in which $\{\Phi_k^{(m+1)}\}$ and $\{\Psi_k^{(n)}\}$ are orthonormal bases in $\mathfrak{F}^{(n+1)}$ and $\mathfrak{F}^{(n)}$, respectively. Since the desired result (3.8) has been established for each term in the series and (see Appendix for notation)

$$\|W(t)\sum_{k=m}^{\infty} |\Phi_k^{(m+1)}\rangle\lambda_k\langle\Psi_k^{(n)}| W^*(t)\|_2^2 \le \sum_{k=1}^{\infty} \lambda_k^2, \tag{3.13}$$

we conclude that (3.8) holds for $P^{(m+1)}AP^{(n)}$ in (3.12). Now, we have already mentioned that (3.8) holds for $m = n = 0$. By taking adjoints in the above relations (3.9)–(3.13) after interchanging the roles of $\Phi^{(m)}$ and $\Psi^{(n)}$, we can prove that if (3.8) holds for some integer values of m and n, then it holds also for m and $n+1$. Hence our induction proof is complete.

Let us consider now the transformers (cf. Appendix)

$$\underline{V}(t)A = V(t)AV^*(t), \quad \underline{U}(t)A = U(t)AU^*(t), \quad (3.14)$$

on $\mathcal{B}_2(\mathfrak{F})$. According to Theorem A1 in the Appendix, $\{\underline{V}(t)\}$ and $\{\underline{U}(t)\}$ are additive one-parameter Abelian groups of unitary transformers in $\mathcal{B}_2(\mathfrak{F})$, with respective generators $\underline{H}$ and $\underline{H}_0$, which can be computed by means of (A18) from H and H_0, respectively. Hence

$$\underline{V}^*(t)A = \underline{V}^{-1}(t)A = V^*(t)AV(t) \quad (3.15)$$

is also unitary in $\mathcal{B}_2(\mathfrak{F})$.

Theorem 3.1: The transformers $\underline{W}(t)$,

$$\underline{W}(t) = \underline{V}^*(t)\underline{U}(t) \quad (3.16)$$

are unitary in $\mathcal{B}_2(\mathfrak{F})$ for any $t \in \mathbb{R}^1$. Furthermore,

$$\underline{W}_\pm A = \operatorname*{w-lim}_{t\to\pm\infty} \underline{P}_\pm \underline{W}(t)A \quad (3.17)$$

for all $A \in \mathcal{B}_2(\mathfrak{F})$, where

$$\underline{W}_\pm A = W_\pm A W_\pm^*, \quad \underline{P}_\pm A = P_\pm A P_\pm. \quad (3.18)$$

Proof: The unitarity of $\underline{W}(t)$ in $\mathcal{B}_2(\mathfrak{F})$ follows directly from the unitarity of $\underline{V}^*(t)$ and $\underline{U}(t)$.

To prove (3.17) note that

$$A = \sum_{m,n=0}^{\infty} P^{(m)}AP^{(n)},$$

where the above series converges to A in a strong sense as well as in the $\|\cdot\|_2$ norm. Write for any Ψ_1, $\Psi_2 \in \mathfrak{F}$

$$\begin{aligned}|\langle \Psi_1 | (\underline{W}_\pm - \underline{P}_\pm \underline{W}(t))A\Psi_2\rangle| &\le \sum_{m+n=0}^{N} |\langle \Psi_1 | (\underline{W}_\pm - \underline{P}_\pm \underline{W}(t))P^{(m)}AP^{(n)}\Psi_2\rangle| \\ &+ \sum_{m+n=N+1}^{\infty} |\langle P^{(m)}W_\pm^*\Psi_1 | AP^{(n)}W_\pm^*\Psi_2\rangle| \\ &+ \sum_{m+n=N+1}^{\infty} \|P^{(m)}W^*(t)\Psi_1\| \, \|P^{(m)}AP^{(n)}W^*(t)\Psi_2\|. \end{aligned} \quad (3.19)$$

The third term in the right-hand side of the above inequality can be majorized uniformly in t by the following expression:

$$\begin{aligned}&\left(\sum_{m+n=N+1}^{\infty} \|P^{(m)}AP^n\|^2\right)^{1/2} \\ &\times\left(\sum_{m+n=N+1}^{\infty} \|P^mW^*(t)\Psi_1\|^2 \|P^{(n)}W^*(t)\Psi_2\|^2\right)^{1/2} \\ &\le \|\Psi_1\| \, \|\Psi_2\| \left(\sum_{m+n=N+1}^{\infty} \|P^{(m)}AP^{(n)}\|_2^2\right)^{1/2}. \end{aligned} \quad (3.20)$$

Since the expression on the right-hand side of (3.20) can be made arbitrarily small by virtue of the fact that $A \in \mathcal{B}_2(\mathfrak{F})$, i.e.,

$$\sum_{m,n=0}^{\infty} \|P^{(m)}AP^{(n)}\|_2^2 = \|A\|_2^2 < \infty,$$

we easily infer from (3.19) and (3.8) that

$$\lim_{t\to\mp\infty}\langle \Psi_1 | (\underline{W}_\pm - \underline{P}_\pm \underline{W}(t))A\Psi_2\rangle = 0.$$

Thus (3.17) is true and Theorem 3.1 is established.

Let $\underline{E}_\lambda$ denote the spectral function of $\underline{H}_0$, for which an explicit formula relating it to E_λ can be found in (A33). Since the generators $\underline{H}$ and $\underline{H}_0$ of $\underline{V}(t)$ and $\underline{U}(t)$, respectively, are self-adjoint in $\mathcal{B}_2(\mathfrak{F})$, the following objects,

$$\underline{W}_{\pm\epsilon} = \underline{P}_\pm \int_{-\infty}^{+\infty} \frac{\mp i\epsilon}{\underline{H} - \lambda \mp i\epsilon} d\underline{E}_\lambda, \quad (3.21)$$

are well-defined transformers in $\mathcal{B}_2(\mathfrak{F})$ for any $\epsilon > 0$.[1]

Theorem 3.2: For any $A \in \mathcal{B}_2(\mathfrak{F})$

$$\operatorname*{w-lim}_{\epsilon\to+0} \underline{W}_{\pm\epsilon}A = \underline{W}_\pm A. \quad (3.22)$$

Proof: An alternative way of writing $\underline{W}_{\pm\epsilon}$ is the following (cf. Ref. 1, p. 436):

$$\underline{W}_{\pm\epsilon} = \mp \epsilon \underline{P}_\pm \int_0^{\mp\infty} \exp(\pm \epsilon t)\underline{W}(t)dt. \quad (3.23)$$

If we set $A_\pm = W_\pm A$, then by using (3.23) we get the following estimate for any $\Psi_1 \in \mathfrak{F}_\pm$, $\Psi_2 \in \mathfrak{F}$:

$$\begin{aligned}|\langle\Psi_1| (\underline{W}_{\pm\epsilon}A - A_\pm)\Psi_2\rangle &\le \mp\epsilon\int_0^{\mp\infty} dt \, \exp(\pm \epsilon t)| \\ &\times \langle \Psi_1 | \Psi \underline{W}(t)_\epsilon A - A_\pm)\Psi_2\rangle|. \end{aligned}$$

Then the rest of the proof is a variation on the proof of the analogous relation (1.4) (cf. Ref. 1, p. 438).

We have to point out that, in computing integral representations for transformers in $\mathcal{B}_2(\mathfrak{F})$ [such as (3.21)] by taking limits of Riemann–Stieltjes sums, these limits could be taken in the Hilbert–Schmidt norm $\|\cdot\|_2$, i.e., they could be h-limits. It is natural then to ask whether the weak limit in (3.22) could not be replaced by an h-limit. The following result shows that this cannot be the case when H_I contains pure creation terms, so that (1.1) does not hold.

Theorem 3.3: In any scattering theory in the Hilbert space $\mathcal{H}$

$$\underline{W}_\pm A = \operatorname*{\mathit{h}-lim}_{t\to\mp\infty} \underline{W}(t)A = \operatorname*{\mathit{h}-lim}_{\epsilon\to+0} \underline{W}_{\pm\epsilon}A, \quad A \in \mathcal{B}_2(\mathcal{H}) \quad (3.24)$$

if and only if $\underline{W}_\pm = \underline{\Omega}_\pm$ satisfy (1.1).

Proof: The first part of the relation (3.24) [which concerns the h-limit of $\underline{W}(t)$] follows immediately from Lemma 6.2 in Ref. 1, p. 334 and Lemma A1, since for $A = |\Psi\rangle\langle\Psi|$

$$\begin{aligned}\lim_{t\to\mp\infty} \|\underline{W}(t)A\|_2^2 &= \lim_{t\to\infty} \|\underline{W}(t)\Psi\|^4 = \|\Psi\|^4 \\ &= \|W_\pm\Psi\|^4 = \|\underline{W}_\pm A\|_2^2. \end{aligned} \quad (3.25)$$

Lemma A1 cannot be applied, however, to $\underline{W}_{\pm\epsilon}$ since $\underline{W}_{\pm\epsilon}$ does not have the form required by that lemma. On the other hand, a direct proof can be given by establishing first that

$$\lim_{\epsilon\to 0}(A | \underline{W}_{\pm\epsilon}B)_2 = (A | \underline{W}_\pm B)_2 \quad (3.26)$$

for arbitrary $A, B \in \mathcal{B}_2(\mathcal{H})$.

Let A be of trace class and have the canonical decomposition (A4), so that

$$\sum_{n=1}^{\infty} \lambda_n = \|A\|_1 < \infty.$$

Then we easily obtain the following estimate:

$$|(A|[\underline{W}_{\pm\epsilon} - \underline{W}_{\pm}]B)_2| = |\sum_{n=1}^{\infty} \langle\Psi_n''|A^*[\underline{W}_{\pm\epsilon} - \underline{W}_{\pm}]B\Psi_n''\rangle| \leq \sum_{n=1}^{N} |\lambda_n\langle\Psi_n'|(\underline{W}_{\pm\epsilon} - \underline{W}_{\pm})B\Psi_n''\rangle| + \|B\|_2 \sum_{n=N+1}^{\infty} \lambda_n, \tag{3.27}$$

in which use has been made of the fact that

$$\|\underline{W}_{\pm\epsilon}B\| \leq \|B\|_2. \tag{3.28}$$

This last inequality easily follows[1] from (3.23).

From the estimate (3.27) we infer that (3.26) holds for $A \in \mathcal{B}_1(\mathcal{H})$. Since $\mathcal{B}_1(\mathcal{H})$ is dense in $\mathcal{B}_2(\mathcal{H})$, we can immediately extend the validity of our conclusion to all $A \in \mathcal{B}_2(\mathcal{H})$ by using (3.28). Thus, we have established that $\underline{W}_{\pm\epsilon}B$ converges to $\underline{W}_{\pm}B$ in the weak topology of the Hilbert space $\mathcal{B}_2(\mathcal{H})$. Since

$$\|\underline{W}_{\pm\epsilon}B\|_2 \leq \|B\|_2 = \lim_{t\to\mp\infty} \|\underline{W}(t)B\|_2 = \|\underline{W}_{\pm}B\|_2, \tag{3.29}$$

we conclude[1] that this convergence is also in convergence in the $\|\cdot\|_2$ norm.

The above theorem shows that we have a scattering theory for density operators $\rho \in \mathcal{B}_1(\mathcal{H}) \subset \mathcal{B}_2(\mathcal{H})$ based on (3.21) (with $\underline{P}_{\pm} = \underline{1}$) and (3.24) at least in the nonrelativistic case when (1.1) holds. In such a theory, instead of working in $\mathcal{H}$, one would work on the Schmidt class $\mathcal{B}_2(\mathcal{H})$, which is itself a Hilbert space. Consequently, from the mathematical point of view such a scattering theory for density operators ρ in $\mathcal{B}_2(\mathcal{H})$ has exactly the same general features as the original theory for state vectors in $\mathcal{H}$. However, this theory presents some obvious advantages from the quantum statistical mechanics point of view.

It should be pointed out that in the general field-theoretical case, when H_I contains pure creation terms and (1.1)–(1.4) do not hold, we still have a time-independent formula for the wave operators $W_{\pm}$:

$$W_{\pm} = \underset{\epsilon\to+0}{\text{w-lim}} P_{\pm} \int_{-\infty}^{+\infty} \frac{\mp i\epsilon}{H - \eta - \lambda \mp i\epsilon}\, dE_{\lambda}. \tag{3.30}$$

The above relation can easily derived from the fact (cf. Ref. 22, Lemma 3.1) that

$$W_{\pm}\Phi_0 = \Phi = \underset{t\to\mp\infty}{\text{w-lim}}\, e^{i(H-\eta)t}e^{-iH_0t} \tag{3.31}$$

(where the phase factor of Φ is chosen that $\langle\Phi|\Phi_0\rangle = 1$), which implies that

$$W_{\pm} = \underset{t\to\pm\infty}{\text{w-lim}} P_{\pm} W(t)e^{-i\eta t}. \tag{3.32}$$

However, by comparison with (3.21)–(3.22) the relation (3.30) has the disadvantage of incorporating the cutoff dependent energy value η (which in realistic models should be the ground-state energy), and therefore it would be less suitable in a renormalization program concerned with the removal of the cutoffs.

4. TIME-INDEPENDENT PERTURBATION THEORY

One of the interesting features of Theorem 3.2 and formula (3.21) is that it opens the possibility of a perturbation computation of the asymptotic states $\rho_{\pm} = \underline{W}_{\pm}\rho$ for any pure or mixed state represented by any trace-class[1,25] statistical operator ρ. The formal expansion leading to a perturbation series can be obtained by applying the standard iterative procedure[1] to the second resolvent equation for $\underline{H}$ and $\underline{H}_0$ to obtain

$$(\underline{H} - \lambda \mp i\epsilon)^{-1} = \underline{R}_0(\lambda \pm i\epsilon) \times\left(\underline{1} - \sum_{n=1}^{\infty} (-1)^{n+1}[\underline{H}_I\underline{R}_0(\lambda \pm i\epsilon)]^n\right), \tag{4.1}$$

where $\underline{R}_0(\zeta)$ is the resolvent of H_0 at ζ and $\underline{H}_I = \underline{H} - \underline{H}_0$. By inserting this expression in (3.21) and noting that

$$\mp \int_{-\infty}^{+\infty} \frac{i\epsilon}{\underline{H}_0 - \lambda \mp i\epsilon}\, d\underline{E}_{\lambda} = \underline{1}, \tag{4.2}$$

we obtain a formal expansion of $\underline{W}_{\pm\epsilon}$ in a series:

$$\underline{W}_{\pm\epsilon} = \underline{P}_{\pm}\left(\underline{1} \pm i\epsilon \sum_{n=1}^{\infty} (-1)^{n+1} \times\int_{-\infty}^{+\infty} \underline{R}_0(\lambda \pm i\epsilon)[\underline{H}_I\underline{R}_0(\lambda \pm i\epsilon)]^n d\underline{E}_{\lambda}\right). \tag{4.3}$$

Of course, the convergence of the above series as well as its behavior when $\epsilon \to +0$ can be investigated only when a more detailed knowledge of H_I, and therefore also of $\underline{H}_I$,

$$\underline{H}_I A = (\underline{H} - \underline{H}_0)A = [H_I, A]_-, \qquad A \in \mathcal{B}_2(\mathcal{F}), \tag{4.4}$$

is available. In carrying out such an investigation it is important to realize that $\underline{E}_{\lambda}$ and $\underline{R}_0(\lambda \pm i\epsilon)$ are known objects, since the formulas (A33) and (A34) relate them to the known quantities E_{λ} and $R_0(\lambda \pm i\epsilon)$, respectively.

The transformers $\underline{W}_{\pm}$ have intertwining properties which are analogous to those of $W_{\pm}$. In fact, from (2.9) we easily obtain

$$W_{\pm} U(s)AU^*(s)W_{\pm}^* = V(s)W_{\pm}AW_{\pm}^*V^*(s) \tag{4.5}$$

for all $A \in \mathcal{B}_2(\mathcal{F})$, i.e.,

$$\underline{W}_{\pm}\underline{U}(s) = \underline{V}(s)\underline{W}_{\pm}. \tag{4.6}$$

If F_{λ} denotes the spectral function of $H - \eta\mathbb{1}$ and $\underline{F}_{\lambda}$ the spectral function of $\underline{H}$, then (4.6) can be[1] recast in the form

$$\underline{W}_{\pm}\underline{E}_{\lambda} = \underline{F}_{\lambda}\underline{W}_{\pm}. \tag{4.7}$$

The above relation implies[1] that

$$\underline{W}_{\pm}\underline{H}_0 A = \underline{H}\, \underline{W}_{\pm} A \tag{4.8}$$

for any $A \in \mathfrak{D}_{\underline{H}_0}$.

In case of groups like $\underline{U}(t)$ and $\underline{V}(t)$ the action of an infinitesimal transformer is related to that of corresponding infinitesimal operators by means of commutators, i.e., by formulas like (A24). Consequently, there is some analogy between the perturbation expansions suggested in (4.3) and the Friedrichs perturbation theory,[26] which adopts Γ-operations involving commutators as its starting point. In fact, Friedrichs approaches the problem of finding a pair of operators $T_{\pm}$ which are such that [cf. (2.9)]

$$T_+H = H_0T_+, \quad HT_- = T_-H_0, \tag{4.9}$$

by setting (cf. Ref. 26, pp. 51–53)

$$T_\pm = \mathbb{1} \pm \Gamma R_\pm. \tag{4.10}$$

Here Γ can be viewed as a transformer in $\mathcal{B}(\mathcal{F})$ which to given $A \in \mathcal{B}(\mathcal{F})$ assigns the solution ΓA of

$$[H_0, \Gamma A] = A. \tag{4.11}$$

If we assume for the moment that $A \in \mathcal{B}_2(\mathcal{F})$, then (4.11) becomes

$$\underline{H}_0 \Gamma A = A, \tag{4.12}$$

i.e., we have $\Gamma = \underline{H}_0^{-1}$ if $\underline{H}_0$ has an inverse (which is actually not the case since, e.g., $\underline{H}_0 A = 0$ if A is a function of H_0).

By taking $R_+ = T_+H_I$ and $R_- = H_I T_-$, Friedrichs arrives (Ref. 26, p. 53) at an iterative procedure leading to the formal expansion

$$T_- = \mathbb{1} - [\Gamma H_I - (\Gamma H_I)^2 + (\Gamma H_I)^4 \mp \cdots].$$

This series can be compared with the formal expansion which is obtained if we take $\epsilon = 0$ in (4.1):

$$\underline{W}_\pm = \underline{P}_\pm \left(\underline{1} \pm \int_{-\infty}^{+\infty} io(H_0 - \lambda)^{-1}\{\underline{H}_I(\underline{H}_0 - \lambda)^{-1} - [\underline{H}_I(\underline{H}_0 - \lambda)^{-1}]^2 \pm \cdots\} d\underline{E}_\lambda.\right.$$

In carrying out such a comparison it should be realized that according to (2.9) and (4.9), T_- plays the role of $W_\pm$, so that we would expect that $\underline{W}_\pm A = T_- A T_-^{-1}$.

ACKNOWLEDGMENTS

The author would like to thank Dr. A. Tip for a very stimulating discussion on the role of transformers in quantum statistical mechanics, and Professor D. Masson for reading the manuscript and contributing critical remarks.

APPENDIX

Let $\mathcal{H}$ be a separable Hilbert space and $\mathcal{B}(\mathcal{H})$ the C^*-algebra of all bounded operators in $\mathcal{H}$. The two-sided ideal $\mathcal{B}_2(\mathcal{H})$ of all Hilbert–Schmidt operators $A \in \mathcal{B}(\mathcal{H})$ becomes a Hilbert space[25] if the inner product $(\cdot|\cdot)_2$ in $\mathcal{B}_2(\mathcal{H})$ is taken to be

$$(A|B)_2 = \mathrm{Tr}(A^*B). \tag{A1}$$

The two-sided ideal $\mathcal{B}_1(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ of all trace class operators A is a dense submanifold of $\mathcal{B}_2(\mathcal{H})$ in the norm $\|\cdot\|_2$ determined by this inner product.

If H_0 is any self-adjoint operator in $\mathcal{H}$ and

$$U(t) = \exp(-iH_0t), \tag{A2}$$

then the mapping

$$A \to \underline{U}(t)A = U(t)AU^*(t) \tag{A3}$$

is a C^*-automorphism of $\mathcal{B}(\mathcal{H})$. We shall refer to linear mappings of submanifolds of $\mathcal{B}(\mathcal{H})$ as *transformers*. Thus, $\underline{U}(t)$ is a norm-preserving transformer on $\mathcal{B}(\mathcal{H})$.

Let us recall[1,25] that every $A \in \mathcal{B}_2(\mathcal{H})$ has a canonical decomposition

$$A = \sum_{n=1}^{\infty} |\Psi_n'\rangle \lambda_n \langle \Psi_n''|, \tag{A4}$$

where $\{\Psi_n'\}$ and $\{\Psi_n''\}$ are orthonormal bases in $\mathcal{H}$ and

$$\sum_{n=1}^{\infty} \lambda_n^2 = \mathrm{Tr}(A^*A) < \infty, \quad \lambda_n \geq 0. \tag{A5}$$

Furthermore, if $A \in \mathcal{B}_1(\mathcal{H})$, then

$$\sum_{n=1}^{\infty} \lambda_n = \|A\|_1 < \infty. \tag{A6}$$

If $A_1, A_2, \cdots$ is any sequence in $\mathcal{B}_2(\mathcal{H})$, we shall write

$$A = h\text{-}\lim_{n\to\infty} A_n \tag{A7}$$

if and only if $A \in \mathcal{B}_2(\mathcal{H})$ and

$$\lim_{n\to\infty} \|A - A_n\|_2 = 0, \tag{A8}$$

where

$$\|A\|_2 = (A|A)_2^{1/2}. \tag{A9}$$

The following lemma relating h-limits to w-limits will prove extremely useful in deriving later results.

Lemma A1: Suppose the bounded operators X_n and Y_n, $n = 1, 2, \cdots$, are such that there is a constant C for which $\|X_n\| \leq C$ and $\|Y_n\| \leq C$ for all $n = 1, 2, \cdots$. If for $\underline{Z}_n A = X_n A Y_n^*$, $A \in \mathcal{B}_2(\mathcal{H})$, there are X, $Y \in \mathcal{B}(\mathcal{H})$ for which

$$\underline{Z}A = XAY^* = \text{w-}\lim_{n\to\infty} \underline{Z}_n A \tag{A10}$$

holds for all $A = |\Psi\rangle\langle\Psi|$, $\Psi \in \mathcal{H}$, and, if for such A

$$\lim_{n\to\infty} \|\underline{Z}_n A\|_2 = \|\underline{Z}A\|_2, \tag{A11}$$

then

$$\underline{Z} = h\text{-}\lim_{n\to\infty} \underline{Z}_n. \tag{A12}$$

Proof: Since any element in $\mathcal{B}_2(A)$ is a linear combination of self-adjoint operators in $\mathcal{B}_2(A)$, it is sufficient to prove the lemma for $A = A^*$. In that case, by the spectral theorem we have

$$A = \sum_{k=1}^{\infty} |\Phi_k\rangle \lambda_k \langle \Phi_k|, \quad \sum_{k=1}^{\infty} \lambda_k^2 = \mathrm{Tr}(A^*A) < \infty, \quad \lambda_k = \overline{\lambda}_k, \tag{A13}$$

where $\{\Phi_k\}$ is an orthonormal basis in $\mathcal{H}$. By virtue of the fact that

$$A = h\text{-}\lim_{m\to\infty} A_m, \quad A_m = \sum_{k=1}^{m} |\Phi_k\rangle \lambda_k \langle \Phi_k|, \tag{A14}$$

and of the following estimate

$$\|(\underline{Z} - \underline{Z}_n)A\|_2 \leq \|(\underline{Z} - \underline{Z}_n)A_m\|_2 + 2C\|A - A_m\|_2, \tag{A15}$$

we can reduce our problem even further, namely to establishing that (A11) holds on the family of all one-dimensional projectors $A = |\Psi\rangle\langle\Psi|$, $\|\Psi\| = 1$.

If $\{\Psi_j\}$ is any orthonormal basis in $\mathcal{H}$, then for the above A

$$\begin{aligned}\|(\underline{Z}-\underline{Z}_n)A\|_2^2 &= \sum_{j=1}^{\infty}\| \,|X\Psi\rangle\langle Y\Psi| - |X_n\Psi\rangle\langle Y_n\Psi|)\Psi_j\|^2 \\ &= \|X\Psi\|^2\|Y\Psi\|^2 + \|X_n\Psi\|^2\|Y_n\Psi\|^2 \\ &\quad - 2\mathrm{Re}\{\langle X_n\Psi|X\Psi\rangle\langle Y\Psi|Y_n\Psi\rangle\}. \end{aligned} \tag{A16}$$

According to (A10) we have

$$\lim_{n\to\infty}\langle X\Psi|X_n\Psi\rangle\langle Y_n\Psi|Y\Psi\rangle = \|X\Psi\|^2\|Y\Psi\|^2,$$

while by (A11)

$$\lim_{n\to\infty}\|X_n\Psi\|^2\|Y_n\Psi\|^2 = \|X\Psi\|^2\|Y\Psi\|^2,$$

and consequently we get from (A16) that (A11) is true.

The following considerations represent a natural application of the ideas on transformers on cross-normed spaces[3-5] to additive one-parameter groups of unitary transformers in $\mathcal{B}_2(\mathcal{H})$.

Lemma A2: For any self-adjoint operators H_1 and H_2 in $\mathcal{H}$ the transformers $\underline{U}_1(t_1)$ and $\underline{U}_2(t_2)$

$$\underline{U}_1(t)A = U_1(t)A, \qquad \underline{U}_2(t)A = AU_2^*(t), \tag{A17}$$

$$U_j(t) = \exp(-iH_jt), \qquad j=1,2 \tag{A18}$$

are unitary operators on $\mathcal{B}_2(\mathcal{H})$ which commute for all $t_1, t_2 \in \mathbb{R}^1$ and which depend h-continuously (i.e., strongly continuously in the $\|\cdot\|_2$-norm) on the parameters t_1 and t_2.

Proof: The commutativity is obvious:

$$\begin{aligned}\underline{U}_1(t_1)\underline{U}_2(t_2)A &= \underline{U}_1(t_1)[AU_2^*(t_2)] = U_1(t_1)AU_2^*(t_2) \\ &= \underline{U}_2(t_2)[U_1(t)A] = \underline{U}_2(t_2)\underline{U}_1(t_1)A.\end{aligned}$$

The unitarity property is equally easy to derive since evidently $\underline{U}_1$ and $\underline{U}_2$ map $\mathcal{B}_2(\mathcal{H})$ onto itself and, e.g.,

$$\begin{aligned}(\underline{U}_2(t)A|\underline{U}_2(t)B)_2 &= \mathrm{Tr}[U_2(t)A^*BU_2^*(t)] \\ &= \mathrm{Tr}(A^*B) = (A|B)_2.\end{aligned}$$

Finally, the h-continuity follows by Lemma A1 from the fact that $\|\underline{U}_j(t)A\|_2 = \|A\|_2$ for all $t\in\mathbb{R}^1$, and from the weak continuity in t of $\underline{U}_j(t)A$ for any $A = |\Psi\rangle\langle\Psi|$. This last property is easily derivable from the weak continuity of $U_j(t)$ since

$$\langle\Psi_1|\underline{U}_1(t)A\Psi_2\rangle = \langle\Psi_1|U_1(t)\Psi\rangle\langle\Psi|\Psi_2\rangle,$$
$$\langle\Psi_1|\underline{U}_2(t)A\Psi_2\rangle = \langle\Psi_1|\Psi\rangle\langle U_2(t)\Psi|\Psi_2\rangle$$

for any $\Psi_1, \Psi_2 \in \mathcal{H}$.

Lemma A3: The families $\{\underline{U}_j(t)\,|\,t\in\mathbb{R}^1\}, j = 1, 2$, of transformers defined in (A17) are additive one-parameter Abelian groups of unitary operators in $\mathcal{B}_2(\mathcal{H})$, whose infinitesimal generators $\underline{H}_j$,

$$\underline{H}_jA = h\text{-}\lim_{t\to 0}(i/t)[\underline{U}_j(t) - \underline{1}]A, \qquad A\in\mathfrak{D}_{\underline{H}_j}, \tag{A19}$$

have the spectral functions

$$\underline{E}_\lambda^{(1)}A = E_\lambda^{(1)}A, \qquad \underline{E}_\lambda^{(2)}A = A(\mathbb{1} - E_{-\lambda}^{(2)}), \tag{A20}$$

where $E_\lambda^{(1)}$ is the spectral function of H_1 and $E_\lambda^{(2)}$ is the spectral function of H_2 modified so as to be continuous from the left.

Proof: Since the derivation for $j=1$ is analogous to that for $j=2$, we shall consider explicitly only the second case, which is slightly more complicated.

The fact that $\underline{U}_2(t)$ is an Abelian additive group is quite obvious:

$$\underline{U}_2(s)\underline{U}_2(t)A = AU_2^*(t)U_2^*(s) = AU_2^*(s+t) = \underline{U}_2(s+t)A.$$

Since $\underline{U}_2(s)$ is h-continuous by Lemma A2, we conclude by Stone's theorem that $\underline{U}_2(s)$ has a generator $\underline{H}_2$ which is a self-adjoint transformer in $\mathcal{B}_2(\mathcal{H})$.

Let us show now that $\underline{E}_\lambda^{(2)}$ is a spectral function in $\mathcal{B}_2(\mathcal{H})$. Since

$$\begin{aligned}\underline{E}_\lambda^{(2)}(\underline{E}_\lambda^{(2)}A) &= A(\mathbb{1}-E_{-\lambda}^{(2)})^2 = A(\mathbb{1}-E_{-\lambda}^{(2)}) = \underline{E}_\lambda^{(2)}A, \\ (\underline{E}_\lambda^{(2)}A|B)_2 &= \mathrm{Tr}[(\mathbb{1}-E_{-\lambda}^{(2)})A^*B] \\ &= \mathrm{Tr}[A^*B(\mathbb{1}-E_{-\lambda}^{(2)})] = (A|\underline{E}_\lambda^{(2)}B)_2,\end{aligned}$$

we conclude that $\underline{E}_\lambda^{(2)}$ is a projector in $\mathcal{B}_2(\mathcal{H})$. Furthermore, evidently $\underline{E}_{-\infty}^{(2)} = 0$, $\underline{E}_\infty^{(2)} = \underline{1}$, and $\underline{E}_\lambda^{(2)} \le \underline{E}_\mu^{(2)}$ for $\lambda\le\mu$ since any $A\in\mathcal{B}_2(\mathcal{H})$ has a canonical decomposition (A4) and therefore

$$\begin{aligned}\|\underline{E}_\lambda^{(2)}A\|_2^2 &= \sum_{n=1}^{\infty}\lambda_n^2|\langle\Psi_n''|(\mathbb{1}-E_{-\lambda}^{(2)})\Psi_n''\rangle|^2 \\ &\le \sum_{n=1}^{\infty}\lambda_n^2|\langle\Psi_n''|(\mathbb{1}-E_{-\mu}^{(2)})\Psi_n''\rangle|^2 \\ &= \|\underline{E}_\mu^{(2)}A\|_2^2.\end{aligned}$$

Finally, $\underline{E}_\lambda^{(2)}$ is continuous from the right, as can be seen if we use (A4) to derive that

$$\|(\underline{E}_\mu^{(2)} - \underline{E}_\lambda^{(2)})A\|_2^2 = \sum_{n=1}^{\infty}\lambda_n^2\|(E_{-\lambda}^{(2)} - E_{-\mu}^{(2)})\Psi_n''\|^2. \tag{A21}$$

In fact, in view of (A5) and of the continuity from the left of $E_\lambda^{(2)}$, we conclude that (A15) converges to zero when $\mu\to\lambda+0$. Thus, $\underline{E}_\lambda^{(2)}$ satisfies all the conditions for a spectral function (cf. Ref. 1, p. 235).

Now we easily compute that

$$\begin{aligned}\int_{-\infty}^{+\infty}e^{-i\lambda t}d\underline{E}_\lambda^{(2)}A &= A\int_{-\infty}^{+\infty}e^{-i\lambda t}d_\lambda(\mathbb{1}-E_{-\lambda}^{(2)}) \\ &= A\exp(iH_2t) = \underline{U}_2(t)A.\end{aligned} \tag{A22}$$

This implies that

$$\underline{H}^{(2)} = \int_{-\infty}^{+\infty}\lambda d\underline{E}_\lambda^{(2)} \tag{A23}$$

since (A22) indicates that $\underline{H}^{(2)}$ and the self-adjoint operator determined by $\underline{E}_\lambda^{(2)}$ are both generators of $\underline{U}_2(t)$, and such generators are unique.

Theorem A1: The family $\{\underline{U}(t)\,|\,t\in\mathbb{R}^1\}$ of transformers (A3) corresponding to a given operator $H_0 = H_0^*$ in $\mathcal{H}$ is an additive one-parameter h-continuous group in $\mathcal{B}_2(\mathcal{H})$. Its generator $\underline{H}_0$ has domain $\mathfrak{D}_{\underline{H}_0}$ containing all $A\in\mathcal{B}_2(\mathcal{H})$, which have a canonical decomposition (A4) with $\Psi_n', \Psi_n'' \in \mathfrak{D}_{H_0}$ and such that $\lambda_n = 0$, $n > N(A)$, for some integer $N(A)$. For any such A

$$\underline{H}_0 A = H_0 A - A H_0. \tag{A24}$$

Proof: The first statement of the theorem follows by Lemma A2 from the fact that

$$\underline{U}(t) = \underline{U}_1(t)\underline{U}_2(t), \qquad H_0 = H_1 = H_2, \tag{A25}$$

and that $U_1(t)$ and $U_2(t)$ commute and are h-continuous. Furthermore, this also tells us that $\underline{H}_0 = \underline{H}_1 + \underline{H}_2$, where $\underline{H}_1$ and $\underline{H}_2$ commute. Consequently

$$\mathfrak{D}_{\underline{H}_0} \supseteq \mathfrak{D}_{\underline{H}_1} \cap \mathfrak{D}_{\underline{H}_2}.$$

Moreover, by Lemma A3, $\mathfrak{D}_{\underline{H}_j}$ contains any A with $\Psi'_n, \Psi''_n \in \mathfrak{D}_{H_0}$ for $n \le N(A)$, and in that case

$$\underline{H}_1 A = H_0 A, \qquad \underline{H}_2 A = - A H_0. \tag{A26}$$

This is so by virtue of the fact that

$$\|t^{-1}[\underline{U}_1(t)A - A]\|_2^2 = \sum_{n=1}^{N(A)} \lambda_n^2 \|t^{-1}[U_1(t)\Psi'_n - \Psi'_n]\|^2,$$

$$\|t^{-1}[\underline{U}_2(t)A - A]\|_2^2 = \sum_{n=1}^{N(A)} \lambda_n^2 \|t^{-1}[U_2^*(t)\Psi''_n - \Psi''_n]\|^2.$$

Thus, Theorem A1 is proven.

Theorem A2: If $f(\lambda)$ is any bounded Borel measurable function on $\mathbb{R}^1$, then

$$f(\underline{H}_0)A = \int_{-\infty}^{+\infty} d_\lambda E_\lambda A f(\lambda - H_0) = \int_{-\infty}^{+\infty} f(H_0 - \lambda)\, A d_\lambda E_\lambda \tag{A27}$$

for any $A \in \mathcal{B}_2(\mathcal{H})$.

Proof: Since the spectral functions $\underline{E}_\lambda^{(1)}$ and $\underline{E}_\lambda^{(2)}$ commute, there is a joint spectral measure $\underline{E}^{(1,2)}(\Delta)$ in $\mathbb{R}^2$ induced by $\underline{E}^{(1)} \times \underline{E}^{(2)}$.[1] In view of the fact that (A25) implies

$$\underline{H}_0 = \underline{H}_1 + \underline{H}_2, \tag{A28}$$

we can write

$$f(\underline{H}_0) = f(\underline{H}_1 + \underline{H}_2) = \int_{\mathbb{R}^2} f(\lambda + \mu) d\underline{E}_{\lambda,\mu}^{(1,2)}. \tag{A29}$$

If we work in the spectral representation space[1] of the commuting transformers H_1, H_2, then using Fubini's theorem applied to $(A|f(\underline{H}_0)B)_2$ for arbitrary $A, B \in \mathcal{B}_2(\mathcal{H})$, we can convert (A29) to either one of the following two relations:

$$f(\underline{H}_0) = \int_{-\infty}^{+\infty} d_\lambda \underline{E}_\lambda^{(1)} \int_{-\infty}^{+\infty} f(\lambda + \mu) d\underline{E}_\mu^{(2)}, \tag{A30}$$

$$f(\underline{H}_0) = \int_{-\infty}^{+\infty} d_\mu \underline{E}_\mu^{(1)} \int_{-\infty}^{+\infty} f(\lambda + \mu) d_\lambda \underline{E}_\lambda^{(2)}. \tag{A31}$$

If we apply now the transformers on both sides of (A30) to any $A \in \mathcal{B}_2(\mathcal{H})$ and employ (A30), we obtain the following integrals with operator-valued integrands[1,6]:

$$f(\underline{H}_0)A = - \int_{-\infty}^{+\infty} d_\lambda E_\lambda \int_{-\infty}^{+\infty} A f(\lambda + \mu) d_\mu E_{-\mu}$$
$$= \int_{-\infty}^{+\infty} d_\lambda E_\lambda A f(\lambda - H_0). \tag{A32}$$

The relation (A31) yields a similar result, thus establishing also the second part of (A27).

Two special cases of (A27) are of special interest. If $f(\mu)$ is the characteristic function of the interval $(-\infty, \lambda]$, then we have $f(\mu - H_0) = 1 - E_{\mu-\lambda}$ and $f(H_0 - \mu) = E_{\lambda-\mu}$,

$$\underline{E}_\lambda A = A - \int_{-\infty}^{+\infty} (d_\mu E_\mu) A E_{\mu-\lambda} = \int_{-\infty}^{+\infty} E_{\lambda-\mu} A d_\mu E_\mu. \tag{A33}$$

This provides a rigorous base for this same relation which has been conjectured in Ref. 27 formula (A25).

For the resolvent $\underline{R}_0(\lambda \mp i\epsilon)$ of $\underline{H}_0$, we obtain

$$(\underline{H}_0 - \lambda \pm i\epsilon)^{-1} A = \int_{-\infty}^{+\infty} (H_0 - \mu + \lambda \mp i\epsilon)^{-1} A d_\mu E_\mu$$
$$= \int_{-\infty}^{+\infty} d_\mu E_\mu A (H_0 - \mu + \lambda \mp i\epsilon)^{-1} \tag{A34}$$

if $\epsilon > 0$ and $A \in \mathcal{B}_2(\mathcal{H})$.

* Supported in part by the National Research Council of Canada.

1 E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971).
2 J. L. Daleckiĭ and S. G. Krein, Amer. Math. Soc. Transl. **47**, 1 (1965).
3 M. S. Birman and M. Z. Solomyak, Sov. Math. **165**, 1567 (1965).
4 M. S. Birman and M. Z. Solomyak, Topics Math. Phys. (Leningrad State Univ.) **1**, 25 (1967); **2**, 19 (1968).
5 M. Z. Solomyak, Topics Math. Phys. (Leningrad State Univ.) **3**, 79 (1969).
6 E. Prugovečki, SIAM J. Math. Anal. **3**, 183 (1972).
7 A. Galindo Tixaire, Helv. Phys. Acta **32**, 412 (1959).
8 M. S. Birman and S. B. Entina, Dokl. Akad. Nauk SSSR **155**, 506 (1964).
9 E. Prugovečki, Nuovo Cimento **63B**, 569 (1969).
10 D. B. Pearson, Nuovo Cimento **2A**, 853 (1971).
11 W. O. Amrein, V. Georgescu, and J. M. Jauch, Helv. Phys. Acta **44**, 407 (1971).
12 W. O. Amrein, P. A. Martin, and B. Misra, Helv. Phys. Acta **43**, 313 (1970).
13 R. B. Lavine, J. Functional Anal. **5**, 368 (1970).
14 V. S. Buslaev and V. B. Matveev, Teor. Mat. Fiz. **2**, 3 (1970).
15 J. V. Corbett, Phys. Rev. D **1**, 3331 (1970).
16 E. Prugovečki, Nuovo Cimento **4B**, 105 (1971).
17 E. Prugovečki, Nuovo Cimento **4B**, 124 (1971).
18 Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **30**, 103 (1963).
19 R. Hoegh-Krohn, Commun. Math. Phys. **21**, 256 (1971).
20 Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **45**, 628 (1971).
21 E. Prugovečki and E. B. Manoukian, Nuovo Cimento B(1972)(to appear).
22 E. Prugovečki and E. B. Manoukian, Commun. Math. Phys. **24**, 133 (1972).
23 R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
24 R. Hoegh-Krohn, J. Math. Phys. **9**, 2075 (1968).
25 T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
26 K. O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (Amer. Math. Soc., Providence, R. I., 1965).
27 A. Tip, Physica **52**, 493 (1971).

On the Equivalence of Dressing Transformations*

Paul Federbush

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104

(Received 6 December 1971; Revised Manuscript Received 22 February 1972)

The equivalence of the representations of the Weyl algebra for ϕ^4_{2+1} in a box induced by the dressing transformation of Glimm and the unitary dressing transformation is studied. Equivalence is shown for the simplified model in which only the most singular portion of the interaction is kept.

I. INTRODUCTION

A method for constructing dressing transformations for certain field theory models has been developed by Glimm.[1] This method employs a sequence of operators that are not unitary. In Refs. 2 and 3 an alternate construction is developed employing unitary operators. In Ref. 4 results are obtained on the equivalence of different dressing transformations constructed for ϕ^4_{2+1} by the method of Glimm. In this paper we obtain some results on the equivalence of the representations of the Weyl algebra induced by the transformations of Glimm and by the unitary transformations for a simplified ϕ^4_{2+1} model—keeping only the most singular terms—in a box.

II. BASIC MODEL AND DEFINITIONS

Corresponding to a cut-off momentum Λ the Hamiltonian is written as

$$H_\Lambda = H_{0\Lambda} + V_\Lambda + \Delta_\Lambda \tag{1}$$

with

$$H_{0\Lambda} = \sum_{|k|\le\Lambda} \omega_k a^*_k a_k, \tag{2}$$

$$\omega_k^2 = k^2 + M^2, \quad V_\Lambda = g\int_0^1 :\phi^4_\Lambda: dx\,|_{0,4+4,0}, \tag{3}$$

$$\Delta_\Lambda = \Delta^{(2)}_\Lambda + \tfrac{1}{2}(\delta M^{(2)}_\Lambda)\int :\phi^2_\Lambda: dx\,|_{1,1}. \tag{4}$$

Here ϕ_Λ contains only operators a_k and a^*_k with $|k| \le \Lambda$: $\Delta^{(2)}_\Lambda$ and $\delta M^{(2)}_\Lambda$ are the renormalization constants appropriate to this modified ϕ^4_{2+1} model, and the numbers following the bars indicate that in the Wick expansion of $:\phi^4_\Lambda:$ only the terms with four creation or four annihilation operators are kept, and in the Wick expansion of $:\phi^2_\Lambda:$ only the terms in one creation and one annihilation operator are kept.

We define V_n as the sum of terms in the expansion of V for which the maximum absolute value of a momentum $k_{\max}$ satisfies

$$(n-1)^\alpha \le k_{\max} < n^\alpha: \tag{5}$$

α is a large number chosen to ensure certain properties of the dressing transformations to be constructed. Note that

$$V = \sum_1^\infty V_n. \tag{6}$$

We use the same definition of Γ as used in Refs. 2 and 3, and let $P(s)$ be the projection onto states with particle number spectra $\le s$.

Define

$$A_n = P(n^2)(\Gamma V_n)P(n^2), \tag{7}$$

$$U_n = e^{-A_n}\cdot e^{-A_{n-1}}\cdots e^{-A_1}. \tag{8}$$

The U_n are unitary operators defining the unitary dressing transformation.

Now let $V^+_n(V^-_n)$ be the terms in V_n containing only creation operators (annihilation operators)

$$V_n = V^+_n + V^-_n. \tag{9}$$

Define

$$B_n = \Gamma V^+_n, \tag{10}$$

$$W_n = \sum_0^{[n/4]} \frac{(-B_n)^s}{s!}, \tag{11}$$

$$T_n = W_n \cdot W_{n-1}\cdots W_1. \tag{12}$$

Equation (11) defines a particular truncation (our construction satisfies condition C_δ of the second paper in Ref. 4). The T_n define Glimm's dressing transformation.

We define $\lambda(n)$:

$$\lambda(n) = \sum_{s=0}^{n} \langle 0|(-\Gamma V^-_s)(\Gamma V^+_s)|0\rangle. \tag{13}$$

Finally we let $\mathfrak{D}$ be the subspace of vectors in Fock space with finite particle number and momentum components in a bounded region of momentum space.

III. GLIMM'S DRESSING TRANSFORMATION

For vectors φ, ψ in $\mathfrak{D}$ we define an inner product

$$\langle\varphi|\psi\rangle_{\mathcal{G}} = \lim_{n\to\infty}\langle T_n\varphi|T_n\psi\rangle\, e^{-\lambda(n)}. \tag{14}$$

Inner products without subscripts indicate inner products in Fock space. The completion of $\mathfrak{D}$ in this product defines a Hilbert space $\mathcal{H}_{\mathcal{G}}$. If G is an appropriate operator—say an element of the Weyl algebra—then the matrix element of the operator G as it appears in this representation is given by

$$\langle\varphi|G|\psi\rangle_{\mathcal{G}} = \lim_{n\to\infty}\langle T_n\varphi|G|T_n\psi\rangle\, e^{-\lambda(n)}. \tag{15}$$

We do not know whether an application of the Gel'fand construction will lead to a Hilbert space larger than $\mathcal{H}_{\mathcal{G}}$.

IV. UNITARY DRESSING TRANSFORMATION

For vectors φ, ψ in $\mathfrak{D}$ we define an inner product

$$\langle\varphi|\psi\rangle_U = \langle\varphi|\psi\rangle. \tag{16}$$

Thus the completion of $\mathfrak{D}$ in this product, $\mathcal{H}_U$, is formally identical to Fock space. If G is an appropriate operator then we shall define a transformed operator G' and define

$$\langle\varphi|G'|\psi\rangle_U = \langle\varphi|G'|\psi\rangle, \tag{17}$$

so that all the information on the unitary dressing transformation is contained in the transformation from G to G'. Formally,

$$G' = \lim_{n\to\infty} U_n^{-1} G U_n. \tag{18}$$

Initially the limit in (18) is defined only for vectors in $\mathfrak{D}$, but since $U_n^{-1}GU_n$ are uniformly bounded by the norm of G, the limit holds on all of $\mathcal{H}_U$. By the methods of Ref. 3 the limit in (18) holds strongly for vectors in $\mathfrak{D}$ and G in the Weyl algebra. The transformation G to G' can clearly be extended to a uniformly closed set of operators.

V. RELATION BETWEEN THE TWO TRANSFORMATIONS

We define the operators R_n:

$$R_n = e^{-(1/2)\lambda(n)} U_n^{-1} T_n, \tag{19}$$

defined initially on $\mathfrak{D}$. We now define

$$R = \lim_{n\to\infty} R_n. \tag{20}$$

We claim R is an isometric operator relating the two dressing transformations as follows:

$$T_{m,n} = W_m \cdot W_{m-1} \cdots W_{n+1}, \tag{21}$$

$$\begin{aligned}\langle\varphi|\psi\rangle_{\mathcal{G}} &= \lim_{m\to\infty} \langle T_m\varphi| T_m\psi\rangle e^{-\lambda(m)}\\ &= \lim_{m\to\infty} \langle U_m^{-1}T_{m,n}U_nR_n\varphi| U_m^{-1}T_{m,n}U_nR_n\psi\rangle e^{-\lambda(m)+\lambda(n)}\\ &= \lim_{n\to\infty}\lim_{m\to\infty} \langle U_m^{-1}T_{m,n}U_nR_n\varphi| U_m^{-1}T_{m,n}U_nR_n\psi\rangle\\ &\quad \times e^{-\lambda(m)+\lambda(n)}\\ &= \lim_{n\to\infty}\langle R_n\varphi|R_n\psi\rangle = \langle R\varphi|R\psi\rangle = \langle R\varphi|R\psi\rangle_U, \end{aligned}\tag{22}$$

$$\begin{aligned}\langle\varphi|G|\psi\rangle &= \lim_{m\to\infty} \langle T_m\varphi|G|T_m\psi\rangle e^{-\lambda(m)}\\ &= \lim_{m\to\infty} \langle U_m^{-1}T_{m,n}U_nR_n\varphi| U_m^{-1}GU_m| U^{-1}T_{m,n}U_nR_n\psi\rangle\\ &\quad\times e^{-\lambda(m)+\lambda(n)}\\ &= \lim_{n\to\infty}\lim_{m\to\infty} \langle U_m^{-1}T_{m,n}U_nR_n\varphi| U_m^{-1}GU_m|\\ &\quad\times U_m^{-1}T_{m,n}U_nR_n\psi\rangle e^{-\lambda(m)+\lambda(n)}\\ &= \langle R\varphi|G'|R\psi\rangle_U. \end{aligned}\tag{23}$$

The statements of (18), (22), and (23) detail the relation between the two dressing procedures:

$$\langle\varphi|\psi\rangle = \langle R\varphi|R\psi\rangle_U, \tag{22$'$}$$

$$\langle\varphi|G|\psi\rangle_{\mathcal{G}} = \langle R\varphi|G'|R\psi\rangle_U. \tag{23$'$}$$

$$G' = \lim_{n\to\infty} U_n^{-1}GU_n, \tag{18$'$}$$

We do not know whether the range of R is all of $\mathcal{H}_U$. The statements of (22) and (23) follow from the following three theorems.

Theorem 1: Let G be an element of the Weyl algebra, $\psi \in \mathfrak{D}$; then

$$\lim_{n\to\infty} U_n^{-1}GU_n|\psi\rangle$$

exists as a strong limit.

Theorem 2: Let $\psi \in \mathfrak{D}$; then

$$\lim_{n\to\infty} R_n|\psi\rangle$$

exists as a strong limit.

Theorem 3: Let $\psi \in \mathfrak{D}$; then

$$\lim_{n\to\infty} \frac{|(U_{m(n)}^{-1}T_{m(n),n}U_n - 1)R_n|\psi\rangle|}{|R_n|\psi\rangle|} = 0,$$

where $m(n)$ is an arbitrary function of n satisfying $m(n) > n$. Theorem 1 may be proved by the methods of Ref. 3, as noted before, much more simply than the result of Ref. 3. We shall prove Theorem 2 and merely comment that Theorem 3 is proved the same way.

VI. PROOF OF THEOREM 2

It is clearly sufficient to prove

$$\sum_{n=m}^{\infty} |(R_n - R_{n-1})|\psi\rangle| < \infty. \tag{24}$$

Now

$$\begin{aligned}R_n - R_{n-1} &= e^{A_1}\cdot e^{A_2}\cdots e^{A_{n-1}}(\Delta_n e^{\Gamma V_n^-} - 1)W_{n-1}\\ &\quad\times W_{n-2}\cdots W_1 e^{-(1/2)\lambda(n-1)}\end{aligned}\tag{25}$$

with

$$\Delta = e^{P\Gamma VP}\sum_0^{[n/4]}\frac{(-\Gamma V^+)^s}{s!}e^{-(1/2)\lambda}e^{-\Gamma V^-}, \tag{26}$$

where to simplify notation we have abbreviated

$$\left.\begin{aligned}\Delta &= \Delta_n,\\ P\Gamma VP &= P(n^2)\Gamma V_n P(n^2),\\ \lambda &= \lambda(n) - \lambda(n-1),\\ V^- &= V_n^-.\end{aligned}\right\}\tag{27}$$

For a given $\psi\in\mathfrak{D}$ and n large enough, the $e^{\Gamma V_n^-}$ in (25) can be dropped so that

$$|(R_n - R_{n-1})|\psi\rangle| = |(\Delta_n - 1)T_{n-1}|\psi\rangle|e^{-(1/2)\lambda(n-1)}. \tag{28}$$

It is known by the work of Glimm (and can also easily be proven using estimate 5 of Ref. 3) that $|T_s|\psi\rangle|e^{-(1/2)\lambda(s)}$ is uniformly bounded in s so that it is sufficient to consider

$$\frac{|(\Delta_n - 1)T_{n-1}|\psi\rangle|e^{-(1/2)\lambda(n-1)}}{|T_{n-1}|\psi\rangle|e^{-(1/2)\lambda(n-1)}} = \delta_n \tag{29}$$

and to show $\sum\delta_n < \infty$. We proceed to study $\Delta - 1$ (using the fact that for n large P may sometimes be replaced by 1):

$$\begin{aligned}\Delta - 1 &= \int_0^1 dt\,\frac{d}{dt}\left(e^{tP\Gamma VP}\sum_0^{[n/4]}\frac{(-t\Gamma V^+)^s}{s!}e^{-(1/2)t^2\lambda}e^{-t\Gamma V^-}\right)\\ &= \int_0^1 dt\,e^{tP\Gamma VP}[C + E_1 + E_2]e^{-t\Gamma V^-}e^{-(1/2)t^2\lambda},\end{aligned}\tag{30}$$

$$E_1 = \Gamma V^+\cdot\frac{(-t\Gamma V^+)^{[n/4]}}{[n/4]!}, \tag{31}$$

$$E_2 = (-t\lambda)\frac{(-t\Gamma V^+)^{[n/4]}}{[n/4]!}, \tag{32}$$

$$\begin{aligned}C &= \sum_0^{[n/4]-1}\frac{(-t\Gamma V^+)^s}{s!}(-t\lambda + [\Gamma V^-, -t\Gamma V^+])\\ &\quad+\sum_{r=2}^{4}\sum_{s=0}^{[n/4]-r}\frac{(-t\Gamma V^+)^s}{s!}\frac{1}{r!}[\Gamma V^-, \underbrace{-t\Gamma V^+,\ldots,-t\Gamma V^+}_{r\text{ terms}}],\end{aligned}\tag{33}$$

where the abbreviations

$$[A,B,C] = [[A,B],C],$$

etc., are used.

On the vectors it acts upon for n large we get

$$|\Delta - 1| \le \int_0^1 dt\, |C + E_1 + E_2| e^{-(1/2)t^2\lambda}. \tag{34}$$

Consider the first term in C

$$\sum_0^{[n/4]-1} \frac{(-t\Gamma V^+)^s}{s!} (-t\lambda + [\Gamma V^-, -t\Gamma V^+]). \tag{35}$$

$|-t\lambda + [\Gamma V^-, -t\Gamma V^+]|$ on the vector upon which it acts may by an N_τ estimate be seen to satisfy

$$|-t\lambda + [\Gamma V^-, -t\Gamma V^+]| < e/N^\beta, \tag{36}$$

where β can be made arbitrarily large by choosing α large. The contribution of this term to $|\Delta - 1|$ contributes a term that for β large ensures $\sum \delta_N < \infty$, provided we can bound $\sum_0^{[n/4]-1} [(-t\Gamma V^+)^s/s!] e^{-(1/2)t^2\lambda}$. The conclusion of the proof is contained in the result that the norm of this term is uniformly bounded in n upon the vectors it acts on. One piece of this is the estimate

$$\sum_0^\infty \frac{x^s}{\sqrt{s!}} \le e' e^{+(1/2)x^2}, \quad x \ge 0, \tag{37}$$

The remaining problem is to estimate $|(\Gamma V_n^+)^s T_{n-1}|\psi\rangle| / |T_{n-1}|\psi\rangle|$.

Assume ψ has particle number restricted by r and all momenta less than or equal to L. If $n > N^{1/\alpha} + 1$, then

$$(\Gamma V_n^-) T_{n-1}|\psi\rangle = 0 \tag{38}$$

and $(\Gamma V^+)^d T_{n-1}|\psi\rangle$ has particle number less than $r + 4d + \sum_{s=0}^{n-1} \frac{1}{4}s \cdot 4$, which is less than $r + 4d + \frac{1}{2}n^2$. For $d \le [\frac{1}{4}n]$ as is restricted in the definition of W_n there is a constant c such that the particle number is $\le cn^2$ for all n. Using the method of the Estimate 5 of Ref. 3, we get for n satisfying the condition preceding (38)

$$\frac{|(\Gamma V^+)^d T_{n-1}|\psi\rangle|}{|T_{n-1}|\psi\rangle|} \le \left[d + \left(\frac{f}{cn^2}\right)^\gamma\right]^{1/2} \times \left[(d-1) + \left(\frac{f}{cn^2}\right)^\gamma\right]^{1/2} \cdots \left[1 + \left(\frac{f}{cn^2}\right)^\gamma\right]^{1/2} \times \lambda^{d/2} \tag{39}$$

for some f and γ, with γ arbitrarily large as α gets large. [In this application of the method of Estimate 5 the grading is not by total number of particles, but by the power of (ΓV^-) that can be applied before the zero vector is obtained.] Estimates (36), (37), and (39)—(36) modified slightly for other terms in $\Delta - 1$—combined yield the theorem.

VII. CONCLUSION

The following is a small selection of questions that remain to be answered. Is the range of R all of $\mathcal{H}_U$? Does the use of the Gel'fand construction lead to an expansion of $H\mathcal{G}$? Is there any way of describing the representation induced by (18)? How unique are the dressing transformations?

---

* This work was supported in part by NSF Grant GP 17523.

1 J. Glimm, Commun. Math. Phys. 5, 343 (1967); 10, 1 (1968); K. Hepp, *Théorie de la renormalization* (Springer, Berlin, 1970).
2 P. Federbush and B. Gidas, Ann. Phys. (N.Y.) 68, 98 (1971); P. Federbush, *ibid.* 68, 94 (1971).
3 P. Federbush, Commun. Math. Phys. 21, 261 (1971).
4 J. Fabrey, Commun. Math. Phys. 19, 1 (1970); J. Eckmann and K. Osterwalder, "On the Uniqueness of the Hamiltonian and of the Representations of the CCR for the Quartic Boson Interaction in Three Dimensions."

A Tachyon Dust Metric in General Relativity

J. C. Foster, Jr. and J. R. Ray
Clemson University, Physics Department, Clemson, South Carolina 29631
(Received 28 July 1971)

Special relativity allows the possibility of a class of particles, called tachyons, which travel with speeds greater than the speed of light in vacuum. These particles have spacelike 4-velocities. Since tachyons have energy and momentum, they will contribute to the gravitational field through the energy–momentum tensor. One question then is what types of solutions to the Einstein field equations will tachyons yield. We consider a metric which admits a four-parameter isometry group. When this metric is used in the field equations using a dust energy–momentum tensor, solutions exist only for spacelike 4-velocity of the dust. We interpret these as solutions for a tachyon dust. Exact solutions to the field equations are obtained.

1. INTRODUCTION

We are interested in solving Einstein's field equations with a dust energy–momentum tensor[1]

$$G^{ij} = R^{ij} - \tfrac{1}{2}g^{ij}R = +\rho U^i U^j + \Lambda g^{ij}, \tag{1.1}$$

where g^{ij} is the metric tensor, R^{ij} the Ricci tensor, R the scalar curvature, Λ the cosmological constant, U^i the 4-velocity of the dust normalized to unity and $\rho > 0$ the energy density. In order to solve the field equations, we assume the space–time possesses symmetries. We shall consider space–times which admit the four-parameter isometry group defined by the commutators[2]

$$[X_1, X_2] = 0, \quad [X_2, X_3] = 0, \quad [X_1, X_3] = X_1, \tag{1.2a}$$

$$[X_1, X_4] = 0, \quad [X_2, X_4] = X_2, \quad [X_3, X_4] = 0. \tag{1.2b}$$

The X_a are the infinitesimal operators of the group and are related to the killing vectors ξ_a^i which generate the group by

$$[A,B,C] = [[A,B],C],$$

etc., are used.

On the vectors it acts upon for n large we get

$$|\Delta - 1| \le \int_0^1 dt\,|C + E_1 + E_2|\,e^{-(1/2)t^2\lambda}. \tag{34}$$

Consider the first term in C

$$\sum_0^{[n/4]-1} \frac{(-t\Gamma V^+)^s}{s!}\,(-t\lambda + [\Gamma V^-, -t\Gamma V^+]). \tag{35}$$

$|-t\lambda + [\Gamma V^-, -t\Gamma V^+]|$ on the vector upon which it acts may by an N_τ estimate be seen to satisfy

$$|-t\lambda + [\Gamma V^-, -t\Gamma V^+]| < e/N^\beta, \tag{36}$$

where β can be made arbitrarily large by choosing α large. The contribution of this term to $|\Delta - 1|$ contributes a term that for β large ensures $\sum \delta_N < \infty$, provided we can bound $\sum_0^{[n/4]-1} [(-t\Gamma V^+)^s/s!]\,e^{-(1/2)t^2\lambda}$. The conclusion of the proof is contained in the result that the norm of this term is uniformly bounded in n upon the vectors it acts on. One piece of this is the estimate

$$\sum_0^\infty \frac{x^s}{\sqrt{s!}} \le e' e^{+(1/2)x^2}, \quad x \ge 0, \tag{37}$$

The remaining problem is to estimate $|(\Gamma V_n^+)^s T_{n-1}|\psi\rangle| / |T_{n-1}|\psi\rangle|$.

Assume ψ has particle number restricted by r and all momenta less than or equal to L. If $n > N^{1/\alpha} + 1$, then

$$(\Gamma V_n^-) T_{n-1}|\psi\rangle = 0 \tag{38}$$

and $(\Gamma V^+)^d T_{n-1}|\psi\rangle$ has particle number less than $r + 4d + \sum_{s=0}^{n-1} \frac{1}{4} s \cdot 4$, which is less than $r + 4d + \frac{1}{2}n^2$. For $d \le [\frac{1}{4}n]$ as is restricted in the definition of W_n there is a constant c such that the particle number is $\le cn^2$ for all n. Using the method of the Estimate 5 of Ref. 3, we get for n satisfying the condition preceding (38)

$$\frac{|(\Gamma V^+)^d T_{n-1}|\psi\rangle|}{|T_{n-1}|\psi\rangle|} \le \left[d + \left(\frac{f}{cn^2}\right)^\gamma\right]^{1/2} \times \left[(d-1) + \left(\frac{f}{cn^2}\right)^\gamma\right]^{1/2} \cdots \left[1 + \left(\frac{f}{cn^2}\right)^\gamma\right]^{1/2} \times \lambda^{d/2} \tag{39}$$

for some f and γ, with γ arbitrarily large as α gets large. [In this application of the method of Estimate 5 the grading is not by total number of particles, but by the power of (ΓV^-) that can be applied before the zero vector is obtained.] Estimates (36), (37), and (39)—(36) modified slightly for other terms in $\Delta - 1$—combined yield the theorem.

VII. CONCLUSION

The following is a small selection of questions that remain to be answered. Is the range of R all of $\mathcal{K}_U$? Does the use of the Gel'fand construction lead to an expansion of $H\mathcal{G}$? Is there any way of describing the representation induced by (18)? How unique are the dressing transformations?

* This work was supported in part by NSF Grant GP 17523.

1 J. Glimm, Commun. Math. Phys. **5**, 343 (1967); **10**, 1 (1968); K. Hepp, *Théorie de la renormalization* (Springer, Berlin, 1970).
2 P. Federbush and B. Gidas, Ann. Phys. (N.Y.) **68**, 98 (1971); P. Federbush, *ibid.* **68**, 94 (1971).
3 P. Federbush, Commun. Math. Phys. **21**, 261 (1971).
4 J. Fabrey, Commun. Math. Phys. **19**, 1 (1970); J. Eckmann and K. Osterwalder, "On the Uniqueness of the Hamiltonian and of the Representations of the CCR for the Quartic Boson Interaction in Three Dimensions."

A Tachyon Dust Metric in General Relativity

J. C. Foster, Jr. and J. R. Ray
Clemson University, Physics Department, Clemson, South Carolina 29631
(Received 28 July 1971)

Special relativity allows the possibility of a class of particles, called tachyons, which travel with speeds greater than the speed of light in vacuum. These particles have spacelike 4-velocities. Since tachyons have energy and momentum, they will contribute to the gravitational field through the energy–momentum tensor. One question then is what types of solutions to the Einstein field equations will tachyons yield. We consider a metric which admits a four-parameter isometry group. When this metric is used in the field equations using a dust energy–momentum tensor, solutions exist only for spacelike 4-velocity of the dust. We interpret these as solutions for a tachyon dust. Exact solutions to the field equations are obtained.

1. INTRODUCTION

We are interested in solving Einstein's field equations with a dust energy–momentum tensor[1]

$$G^{ij} = R^{ij} - \tfrac{1}{2}g^{ij}R = +\rho U^i U^j + \Lambda g^{ij}, \tag{1.1}$$

where g^{ij} is the metric tensor, R^{ij} the Ricci tensor, R the scalar curvature, Λ the cosmological constant, U^i the 4-velocity of the dust normalized to unity and $\rho > 0$ the energy density. In order to solve the field equations, we assume the space–time possesses symmetries. We shall consider space–times which admit the four-parameter isometry group defined by the commutators[2]

$$[X_1, X_2] = 0, \quad [X_2, X_3] = 0, \quad [X_1, X_3] = X_1, \tag{1.2a}$$

$$[X_1, X_4] = 0, \quad [X_2, X_4] = X_2, \quad [X_3, X_4] = 0. \tag{1.2b}$$

The X_a are the infinitesimal operators of the group and are related to the killing vectors ξ_a^i which generate the group by

$$X_a = \xi_a^i \frac{\partial}{\partial x^i} = \xi_a^i P_i. \tag{1.3}$$

We suppose that the four-parameter group acts on three-dimensional surfaces of transitivity. Then from the form of the commutators, Eqs. (1.2), it follows that the metric and infinitesimal operators can be written[3]

$$ds^2 = A(x^3)(dx^1)^2 + 2C(x^3)e^{+x^1}dx^2dx^4 + (dx^3)^2, \tag{1.4}$$

$$X_1 = P_2, \quad X_2 = P_4, \quad X_3 = -P_1 + x^2P_2, \quad X_4 = -P_1 + x^4P_4. \tag{1.5}$$

Here we have employed a semigeodesic coordinate system with geodesically parallel hypersurfaces $x^3 =$ const. From the field equations (1.1) we find

$$\Delta_{\xi_a} U^i = 0, \tag{1.6}$$

where Δ_{ξ_a} denotes the Lie derivative with respect to the Killing vector ξ_a^i. Equations (1.6) imply that U^i has the form

$$U^i = (\alpha(x^3), 0, \beta(x^3), 0). \tag{1.7}$$

The following coordinate transformation diagonalizes the metric:

$$x'^2 = 2^{-1/2}(x^2 + x^4), \quad x'^4 = 2^{-1/2}(-x^2 + x^4).$$

In the new coordinates, after dropping the primes,

$$ds^2 = A(x^3)(dx^1)^2 + C(x^3)e^{+x^1}(dx^2)^2 + (dx^3)^2 - C(x^3)e^{+x^1}(dx^4)^2, \tag{1.8}$$

$$U^i = (\alpha(x^3), 0, \beta(x^3), 0). \tag{1.9}$$

In these coordinates the determinant of the metric tensor g_{ik} is

$$g = -AC^2e^{+2x^1}.$$

Therefore, in order to have an acceptable solution, we must require $A > 0$. This means, depending on the sign of C, x^2 or x^4 must be the timelike coordinate. We choose $C > 0$ and therefore x^4 is the timelike coordinate. The other choice gives the same final results. Since x^4 is the timelike coordinate, it follows from Eqs. (1.9) that our space–time allows dust solutions only for spacelike U^i. We interpret these solutions as solutions to the Einstein field equations for a tachyon dust.[4,5] With this interpretation ρ is the invariant momentum flux of the tachyon dust.

2. THE FIELD EQUATIONS

Since U^i is spacelike,

$$U^iU_i = 1. \tag{2.1}$$

We assume that the tachyon dust moves along geodesics

$$\frac{DU^i}{ds} = U^i_{;j}U^j = 0, \tag{2.2}$$

where the semicolon denotes covariant differentiation. The solution of Eqs. (2.1), (2.2) using Eqs. (1.8), (1.9) yields

$$A(x^3)\alpha(x^3) = C = \text{const}, \quad \alpha \neq 0. \tag{2.3}$$

Next we perform the coordinate transformation[6]

$$x'^1 = \int (\alpha/\beta)\,dx^3 - x^1, \tag{2.4a}$$

$$x'^3 = \int \beta dx^3 + Cx^1, \quad \beta \neq 0. \tag{2.4b}$$

In the new coordinates, after dropping the primes,

$$ds^2 = X^2(x^1,x^3)(dx^1)^2 + Y^2(x^1,x^3)(dx^2)^2 + (dx^3)^2 - Y^2(x^1,x^3)(dx^4)^2, \tag{2.5}$$

$$U^i = (0,0,1,0). \tag{2.6}$$

The infinitesimal operators in this coordinate system are

$$X_1 = 2^{-1/2}(P_2 - P_4), \tag{2.7a}$$

$$X_2 = 2^{-1/2}(P_2 + P_4), \tag{2.7b}$$

$$X^3 = P_1 + 2^{-1}(x^2 - x^4)(P_2 - P_4) - CP_3, \tag{2.7c}$$

$$X_4 = P_1 + 2^{-1}(x^2 + x^4)(P_2 + P_4) - CP_3. \tag{2.7d}$$

Applying Killing's equations,

$$\Delta_{\xi_a} g_{ik} = 0, \tag{2.8}$$

in the new coordinates yields the following equations for the metric:

$$X_1 - CX_3 = 0, \tag{2.9}$$

$$Y_1 - CY_3 + 2^{-1}Y = 0, \tag{2.10}$$

where we have introduced the notation

$$X_\alpha = \frac{\partial X}{\partial x^\alpha}, \quad Y_\alpha = \frac{\partial Y}{\partial x^\alpha}, \quad \alpha = 1, 3.$$

Equation (2.9) yields

$$X = h(Cx^1 + x^3), \tag{2.11}$$

where h is an arbitrary function. The field equations (1.1) become

$$2\frac{Y_{33}}{Y} + \frac{Y_3^2}{Y^2} + \frac{Y_1^2}{X^2Y^2} = \Lambda, \quad i,j = 1,1, \tag{2.12a}$$

$$\frac{Y_{33}}{Y} + \frac{X_{33}}{X} + \frac{X_3Y_3}{XY} + \frac{1}{X^2}\left(\frac{Y_{11}}{Y} - \frac{X_1Y_1}{XY}\right) = \Lambda, \quad i,j = 2,2 \text{ or } 4,4, \tag{2.12b}$$

$$\frac{Y_3^2}{Y^2} + 2\frac{X_3Y_3}{XY} + \frac{1}{X^2}\left(2\frac{Y_{11}}{Y} + \frac{Y_1^2}{Y^2} - 2\frac{X_1Y_1}{XY}\right) = \rho + \Lambda, \quad i,j = 3,3, \tag{2.12c}$$

$$\frac{Y_{13}}{Y} - \frac{X_3Y_1}{XY} = 0, \quad i,j = 1,3 \text{ or } 3,1. \tag{2.12d}$$

Next we present explicit solutions to the field equations. Our equations and solutions are very similar to those considered by Farnsworth in his investigation of metrics allowing a G_4V group of isometries.[6]

3. SOLUTIONS FOR $Y_1 = 0$

If $Y_1 = 0$, Eq. (2.10) yields

$$Y = Ae^{x^3/2C}, \quad A = \text{const}. \tag{3.1}$$

Using Eqs. (2.11), (3.1) in the field equation (2.12a) gives

$$\Lambda = 3/4C^2. \tag{3.2}$$

Therefore, the $Y_1 = 0$ solutions must have a positive cosmological constant. Using Eqs. (2.11), (3.1), (3.2) in the field equation (2.12b) yields the solution

$$X(x^1, x^3) = Be^{\eta/2} - De^{-\eta}, \quad B, D = \text{constants}, \quad \eta = x^1 + x^3/C. \tag{3.3}$$

Finally using Eqs. (2.11), (3.1), (3.2), (3.3) in field equation (2.12c) yields

$$\rho = (3/2C^2)\left[(B/D)\, e^{3\eta/2} - 1\right]^{-1} \tag{3.4}$$

for the momentum flux. The space-time is singular on the hypersurface

$$\eta = \tfrac{2}{3}\ln(D/B). \tag{3.5}$$

The momentum flux is shown in Fig. 1 for a particular choice of B, C, and D.

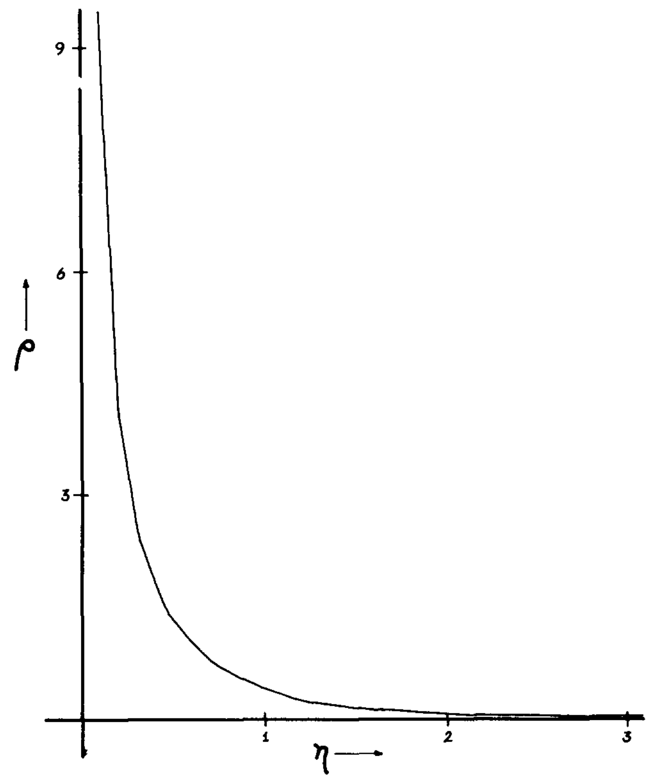

FIG. 1. Momentum flux of tachyon dust for the case $Y_1 = 0$, $B = C = D = 1$.

4. SOLUTIONS FOR $Y_1 \neq 0$

For $Y_1 \neq 0$, field equation (2.12d) yields

$$X = G(x^1)Y_1, \tag{4.1}$$

where G is an arbitrary function. Killing's equation (2.10) yields

$$Y = g(Cx^1 + x^3)e^{-x^1/2}, \tag{4.2}$$

where g is an arbitrary function. Combining Eqs. (2.11), (4.1), (4.2) yields

$$G(x^1) = Ee^{x^1/2}, \quad E = \text{const},$$

and therefore by Eqs. (4.1), (4.2)

$$X = E(Cg' - \tfrac{1}{2}g), \tag{4.3}$$

where

$$g' = dg(Cx^1 + x^3)/d(Cx^1 + x^3).$$

The remaining field equations yield

$$2gg'' + g'^2 + E^{-2} - \Lambda g^2 = 0, \tag{4.4}$$

$$\rho = (3g'' - \Lambda g)/(Cg' - \tfrac{1}{2}g). \tag{4.5}$$

For $\Lambda = 0$ the derivative of Eq. (4.4) yields

$$g^2 g'' = -F = \text{const},$$

which when combined with Eq. (4.4) yields

$$g'^2 = -1/E^2 + 2F/g, \tag{4.6}$$

the Friedmann equation. For $C = E = F/2 = 1$ this yields the solution

$$g = 2(1 - \cos\psi), \quad \eta = 2(\psi - \sin\psi),$$
$$\rho = 3/2(1 - \cos\psi)[(1 - \cos\psi)^2 - \sin\psi]. \tag{4.7}$$

This solution also has singularities on hypersurfaces $\eta = \text{const}$. The momentum flux is shown in Fig. 2 for this solution.

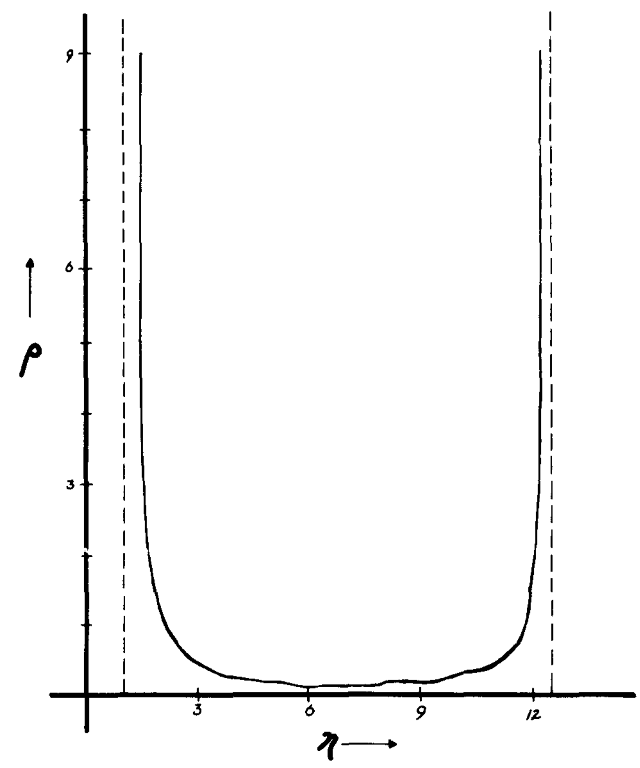

FIG. 2. Momentum flux of tachyon dust for the case $Y_1 \neq 0$, $\Lambda = 0$, $C = E = F/2 = 1$.

5. CONCLUSIONS

We have discussed a metric which allows a four-parameter isometry group G_4IV. When this metric is used in Einstein's field equations with a dust energy–momentum tensor, one finds solutions only for a tachyon dust. The solutions are static and have singularities on timelike hypersurfaces. Our solutions do not, of course, prove anything concerning the existence or nonexistence of tachyon dust in nature. What they do show is that the Einstein field equations have solutions for such a hypothetical dust. In fact the metric we considered has solutions only for this type of dust.

ACKNOWLEDGMENTS

One of us (J.R.R.) expresses his appreciation to the Clemson University Faculty Research Council for a grant which made possible the presentation of an abbreviated form of this work at the International Conference on Gravitation and Relativity held in Copenhagen, July 1971.

1 $i, j, k \cdots$ are tensor indices which take on the values 1, 2, 3, 4. $a, b, c, \cdots$ label the Killing vectors, infinitesimal operators, etc. For a four-parametric group these also take on the values 1, 2, 3, 4. We choose the signature (+, +, +, −). Also we use units with $c = 1$, $8\pi k = 1$, where c is the speed of light in vacuum and k is the Newtonian gravitational constant.

2 A. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
3 See Ref. 2, p. 229. Our case is classified as G_4IV in Ref. 2.
4 O. Bilaniuk, V. Deshpande, and E. Sudarshan, Amer. J. Phys. **30**, 718 (1962).
5 G. Feinberg, Phys. Rev. **159**, 1089 (1967).
6 D. Farnsworth, J. Math. Phys. **8**, 2315 (1967).

Time-Dependent and Time-Independent Potential Scattering for Asymptotically Coulomb Potentials

Dan Marchesin and Michael L. O'Carroll
Department of Mathematics, Pontificia Universidade Católica, R. Marques de Sao Vicente, 209, Rio de Janeiro, Brasil

(Received 27 August 1971; Revised Manuscript Received 7 February 1972)

We consider the three-dimensional quantum mechanical problem of nonrelativistic potential scattering, where the Hamiltonian is $H_2 = H_1 + V = H_0 + V_c + V$; H_0 is the free particle Hamiltonian, V_c is the Coulomb potential, and $V(x)$ is a real-valued potential function defined for $x \in R^3$. We show the existence of the wave operators $W_\pm(H_2, H_D) = \text{s-lim}\, e^{iH_2 t} e^{-iH_D(t)}$ on $L^2(R^3)$ for $V = V_2 + V'$, where $V_2 \in L^2(R^3)$, $V' \in L^\infty(R^3) \cap L^p(R^3)$ $(p < 3)$, and $e^{-iH_D(t)}$ is the family of unitary operators used by Dollard to show the existence of the wave operators $W_\pm(H_1, H_D)$ appropriate for the pure Coulomb case. If V is spherically symmetric then $W_\pm(H_2, H_D)$ are shown to be absolutely continuous complete. If in addition $V(r)$ is continuous in $(0, \infty)$, then $W_\pm(H_2, H_D)$ are continuous complete. In both cases $S = W_+^* W_-$ is unitary. The connection between the more physical, time-dependent wave operator approach and the traditional time-independent method is made, and phase shift formulas are obtained for the wave and scattering operators.

INTRODUCTION

One of the first problems to be explicitly solved in nonrelativistic quantum mechanics was the Coulomb scattering problem by the time-independent eigenfunction expansion method.[1] Scattering solutions are obtained by imposing certain asymptotic conditions on the solutions $\Phi_1^\pm(x, q)$ of the partial differential equation $(-\Delta + V_c - k^2)\Phi_1^\pm = 0$ (see Appendix A). Throughout this paper, x and q will be used to denote elements of R^3 while r and k will denote their magnitudes.

Let us consider the three-dimensional nonrelativistic system described by the Hamiltonian $H_2 = H_0 + V_c + V = H_1 + V$, where H_0 is the free particle Hamiltonian, V_c is the Coulomb potential, and V is a real-valued potential function. H_i $(i = 0, 1, 2)$ act in the Hilbert space $\mathcal{H} = L^2(R^3)$. Recently the existence of certain modified wave operators

$$W_\pm(H_2, H_D) = \underset{t \to \pm\infty}{\text{s-lim}}\, e^{iH_2 t} e^{-iH_D(t)}$$

appropriate for systems where the potential function is asymptotically Coulomb was shown by Dollard,[2] thus establishing a more physical time-dependent description of the scattering process. Furthermore, when the potential is pure Coulomb, Dollard[2] showed the absolutely continuous spectrum completeness (also continuous spectrum completeness) of the wave operators $W_\pm(H_1, H_D)$ and established the relation between the customary scattering solutions of the time-independent theory. Support for the acceptance of the wave operators $W_\pm(H_1, H_D)$ as the appropriate ones for the description of the pure Coulomb scattering process is given by the following facts[2,3]:

(1) Let P_1 be the projection on the continuous spectrum subspace of H_1. For any vector $f \in P_1\mathcal{H}$ there exists vectors $g_\pm \in \mathcal{H}$ such that

$$\lim_{t \to \pm\infty} \| e^{-iH_1 t} f_\pm - e^{-iH_D(t)} g_\pm \| = 0.$$

Conversely, for every $g \in \mathcal{H}$ there exist vectors $f_\pm \in \mathcal{H}$ such that

$$\lim_{t \to \pm\infty} \| e^{-iH_1 t} f_\pm - e^{-iH_D(t)} g \| = 0.$$

This implies the existence of the wave operators $W_\pm(H_1, H_D)$.

Furthermore

$$\lim_{t \to \pm\infty} \int_B | e^{-iH_1 t} f(x) |^2 dx = 0$$

for any compact subset B of R^3, so that indeed $e^{-H_1 t} f$ behaves as a traveling wavepacket.

(2) The momentum and position probability distributions of the vector $e^{-iH_D(t)} g_+$ $(e^{-iH_D(t)} g_-)$ in the $t \to +\infty$ $(t \to -\infty)$ limit are the same as those of the vector $e^{-iH_0 t} g_+$ $(e^{-iH_0 t} g_-)$.

The momentum distribution of $e^{-iH_0 t} h$ is just $|h_0(k)|^2$ and the position probability distribution is given by $(m/t)^3 |h_0(mx/t)|^2$ as t approaches $\pm\infty$ (h_0 is the three-dimensional Fourier transform of h).

5. CONCLUSIONS

We have discussed a metric which allows a four-parameter isometry group G_4IV. When this metric is used in Einstein's field equations with a dust energy–momentum tensor, one finds solutions only for a tachyon dust. The solutions are static and have singularities on timelike hypersurfaces. Our solutions do not, of course, prove anything concerning the existence or nonexistence of tachyon dust in nature. What they do show is that the Einstein field equations have solutions for such a hypothetical dust. In fact the metric we considered has solutions only for this type of dust.

ACKNOWLEDGMENTS

One of us (J.R.R.) expresses his appreciation to the Clemson University Faculty Research Council for a grant which made possible the presentation of an abbreviated form of this work at the International Conference on Gravitation and Relativity held in Copenhagen, July 1971.

1 $i, j, k \cdots$ are tensor indices which take on the values 1, 2, 3, 4. $a, b, c, \cdots$ label the Killing vectors, infinitesimal operators, etc. For a four-parametric group these also take on the values 1, 2, 3, 4. We choose the signature (+, +, +, −). Also we use units with $c = 1$, $8\pi k = 1$, where c is the speed of light in vacuum and k is the Newtonian gravitational constant.

2 A. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
3 See Ref. 2, p. 229. Our case is classified as G_4IV in Ref. 2.
4 O. Bilaniuk, V. Deshpande, and E. Sudarshan, Amer. J. Phys. **30**, 718 (1962).
5 G. Feinberg, Phys. Rev. **159**, 1089 (1967).
6 D. Farnsworth, J. Math. Phys. **8**, 2315 (1967).

Time-Dependent and Time-Independent Potential Scattering for Asymptotically Coulomb Potentials

Dan Marchesin and Michael L. O'Carroll

Department of Mathematics, Pontificia Universidade Católica, R. Marques de Sao Vicente, 209, Rio de Janeiro, Brasil

(Received 27 August 1971; Revised Manuscript Received 7 February 1972)

We consider the three-dimensional quantum mechanical problem of nonrelativistic potential scattering, where the Hamiltonian is $H_2 = H_1 + V = H_0 + V_c + V$; H_0 is the free particle Hamiltonian, V_c is the Coulomb potential, and $V(x)$ is a real-valued potential function defined for $x \in R^3$. We show the existence of the wave operators $W_\pm(H_2, H_D) = \text{s-lim}\, e^{iH_2 t} e^{-iH_D(t)}$ on $L^2(R^3)$ for $V = V_2 + V'$, where $V_2 \in L^2(R^3)$, $V' \in L^\infty(R^3) \cap L^p(R^3)$ $(p < 3)$, and $e^{-iH_D(t)}$ is the family of unitary operators used by Dollard to show the existence of the wave operators $W_\pm(H_1, H_D)$ appropriate for the pure Coulomb case. If V is spherically symmetric then $W_\pm(H_2, H_D)$ are shown to be absolutely continuous complete. If in addition $V(r)$ is continuous in $(0, \infty)$, then $W_\pm(H_2, H_D)$ are continuous complete. In both cases $S = W_+^* W_-$ is unitary. The connection between the more physical, time-dependent wave operator approach and the traditional time-independent method is made, and phase shift formulas are obtained for the wave and scattering operators.

INTRODUCTION

One of the first problems to be explicitly solved in nonrelativistic quantum mechanics was the Coulomb scattering problem by the time-independent eigenfunction expansion method.[1] Scattering solutions are obtained by imposing certain asymptotic conditions on the solutions $\Phi_1^\pm(x, q)$ of the partial differential equation $(-\Delta + V_c - k^2)\Phi_1^\pm = 0$ (see Appendix A). Throughout this paper, x and q will be used to denote elements of R^3 while r and k will denote their magnitudes.

Let us consider the three-dimensional nonrelativistic system described by the Hamiltonian $H_2 = H_0 + V_c + V = H_1 + V$, where H_0 is the free particle Hamiltonian, V_c is the Coulomb potential, and V is a real-valued potential function. H_i $(i = 0, 1, 2)$ act in the Hilbert space $\mathcal{H} = L^2(R^3)$. Recently the existence of certain modified wave operators

$$W_\pm(H_2, H_D) = \underset{t \to \pm\infty}{\text{s-lim}}\, e^{iH_2 t} e^{-iH_D(t)}$$

appropriate for systems where the potential function is asymptotically Coulomb was shown by Dollard,[2] thus establishing a more physical time-dependent description of the scattering process. Furthermore, when the potential is pure Coulomb, Dollard[2] showed the absolutely continuous spectrum completeness (also continuous spectrum completeness) of the wave operators $W_\pm(H_1, H_D)$ and established the relation between the customary scattering solutions of the time-independent theory. Support for the acceptance of the wave operators $W_\pm(H_1, H_D)$ as the appropriate ones for the description of the pure Coulomb scattering process is given by the following facts[2,3]:

(1) Let P_1 be the projection on the continuous spectrum subspace of H_1. For any vector $f \in P_1\mathcal{H}$ there exists vectors $g_\pm \in \mathcal{H}$ such that

$$\lim_{t \to \pm\infty} \| e^{-iH_1 t} f_\pm - e^{-iH_D(t)} g_\pm \| = 0.$$

Conversely, for every $g \in \mathcal{H}$ there exist vectors $f_\pm \in \mathcal{H}$ such that

$$\lim_{t \to \pm\infty} \| e^{-iH_1 t} f_\pm - e^{-iH_D(t)} g \| = 0.$$

This implies the existence of the wave operators $W_\pm(H_1, H_D)$.

Furthermore

$$\lim_{t \to \pm\infty} \int_B | e^{-iH_1 t} f(x) |^2 dx = 0$$

for any compact subset B of R^3, so that indeed $e^{-H_1 t} f$ behaves as a traveling wavepacket.

(2) The momentum and position probability distributions of the vector $e^{-iH_D(t)} g_+$ $(e^{-iH_D(t)} g_-)$ in the $t \to +\infty$ $(t \to -\infty)$ limit are the same as those of the vector $e^{-iH_0 t} g_+$ $(e^{-iH_0 t} g_-)$.

The momentum distribution of $e^{-iH_0 t} h$ is just $|h_0(k)|^2$ and the position probability distribution is given by $(m/t)^3 |h_0(mx/t)|^2$ as t approaches $\pm\infty$ (h_0 is the three-dimensional Fourier transform of h).

(3) $(W_\pm g)(x) = \int \Phi_1^\pm(x,q) g_0(q)dq = f_\pm(x)$ and the time evolution $e^{-iH_1 t} f_\pm$ is given by $f_\pm(x,t) = \int \Phi_1^\pm(x,q) \times e^{-ik^2 t} g_0(q)dq$ thus allowing the identification of the variable q occurring in $\Phi_1^\pm(x,q)$ as the asymptotic $(t \to \pm\infty)$ momentum index.

In this article we obtain results similar to those of Dollard for a system described by the Hamiltonian H_2. In Sec. 2 we extend Dollard's existence proof for $W_\pm(H_2, H_D)$ to include a larger class of potentials (not necessarily spherically symmetric), which include potentials behaving as $r^{-1-\epsilon}$ as $r \to \infty$ $(\epsilon > 0)$. In the case when V is spherically symmetric, we show absolutely continuous completeness of the wave operators $W_\pm(H_2, H_D)$. Since an eigenfunction expansion has not been shown to exist for H_2 for nonspherically symmetric potentials, we use Kodaira's theory of eigenfunction expansions[4] for spherically symmetric potentials and consider direct sums over the angular momentum to establish the relation between the wave operators $W_\pm(H_2, H_D)$ and the time-independent method.

The notation we will use is introduced in Sec. 1. Some facts from the theory of eigenfunction expansion are reviewed there also. In Sec. 2 the theorems are presented: Their proofs follow in Sec. 3. Appendix A gives formulas connecting three- and one-dimensional expressions. In Appendix B we give a heuristic argument, for spherically symmetric V, which indicates that the wave operators $W_\pm(H_2, H_D)$ exist and are absolutely continuous complete when V satisfies Kodaira's criterion [see Eqs. (1a)—(1c)] for the existence of an eigenfunction expansion for H_2.

For another approach to this problem the reader is referred to the article by Mulherin and Zinnes.[5] For the treatment of the existence and absolutely continuous completeness question, without the spherically symmetric restriction, see forthcoming work of B. Simon. For the case of systems described by potentials of the type c/r^α $(0 < \alpha < 1)$ see Amrein, Martin, and Misra.[6] We also mention that results analogous to ours have been obtained by Green and Lanford[7] (see also Kuroda[8]) for the case $V_c = 0$, $H_2 = H_0 + V$.

1. PRELIMINARIES

Throughout this paper x, q denote elements of $\mathbb{R}^3$ while r, k will denote their magnitudes. We adhere to the Hilbert space terminology of Kato.[9] We consider the formally symmetric operators H_i $(i = 0,1,2)$, where H_0 is the kinetic energy operator, H_1 is $H_0 + V_c$ ($V_c = e_1 e_2 / r$, the Coulomb potential), and $H_2 = H_1 + V$. V is a real-valued measurable function defined for $x \in \mathbb{R}^3$. The operators H_i act in $\mathcal{L}^2(\mathbb{R}^3)$ and their definition as self-adjoint operators will be obtained either from a form-bounded or operator relative-boundedness condition. We will not change notation for the corresponding self-adjoint extensions of these operators. When we are dealing with a spherically symmetric potential V we can also consider these operators as direct sums, i.e., $H_i = \sum_{l=0}^\infty \oplus H_i^l$ $(i = 0,1,2)$, which act in the direct sum of the Hilbert spaces $\mathcal{H}^l$. $\mathcal{H} = \sum_{l=0}^\infty \oplus \mathcal{H}^l$ is isomorphic to $\mathcal{L}^2(\mathbb{R}^3)$ (see Green and Lanford[7]). The subspace $\mathcal{H}^l$ is the Hilbert space $\mathcal{H}^l = \mathcal{L}^2(0,\infty) \otimes \mathcal{L}_l^2(\Omega)$, where $\mathcal{L}_l^2(\Omega)$ is the $(2l+1)$-dimensional Hilbert space with orthonormal basis elements given by the spherical harmonics $Y_{lm}(\Omega)$, $-l \le m \le l$, and $\mathcal{L}^2(0,\infty)$ is the radial Hilbert space with measure dr. H_i^l $(i = 0,1,2)$ acts trivially on the second factor. The definition of H_i^l $(i = 0,1,2)$ as a self-adjoint operator in $\mathcal{L}^2(0,\infty)$ will be taken from the theory of eigenfunction expansions as developed by Kodaira[4] which we will briefly review. These operators are self-adjoint restrictions of the differential operators

$$\mathcal{L}_i^l = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V_i'(r), \qquad i = 0,1,2,$$

where $V_0' = 0$, $V_1' = V_c$, and $V_2' = V_c = V_c + V$. For $l = 0$ we are in the limit circle case so we impose the boundary condition $u(0) = 0$ for $u \in D(H_i^0)$. For the precise specification of the domains of these operators see Kodaira[4] or Stone.[10] When V satisfies the conditions

$$V(r) \text{ continuous on each closed subinterval of } (0,\infty) \tag{1.1a}$$

$$V(r) \sim O(1/r^{2-\epsilon}) \quad \text{for } r \to 0,\ \epsilon > 0, \tag{1.1b}$$

$$V(r) \sim O(1/r^{1+\epsilon'}) \quad \text{for } r \to \infty,\ \epsilon' > 0, \tag{1.1c}$$

then H_i^l $(i = 0,1,2)$ admit eigenfunction expansions, i.e., for any $f \in \mathcal{L}^2(0,\infty)$ (we suppress the l dependence),

$$f(r) = \text{l.i.m.} \sum_{n=1}^{N} u_{1n}(r)(u_{1n}, f) + \underset{N\to\infty}{\text{l.i.m.}} \int_{N^{-1}}^{N} u_i(r,k) \int_0^\infty u_i(r,k) f(r)\,dr\ dk,$$

where u_{1n} are the orthonormal eigenvectors of the point spectrum of H_i^l and $u_i(r,k)$ is a real continuous function of r $(0 \le r < \infty)$ and k $(k > 0)$, which satisfies

$$\begin{aligned} &(L_i^l - k^2) u_i^l(r,k) = 0, \qquad k > 0, \\ &u_i^l(r,k) \to C(k) r^{l+1}, \qquad \text{as } r \to 0,\ k > 0, \end{aligned} \tag{1.2a}$$

$$u_0(r,k) = (2/\pi)^{1/2} kr\, j_l(kr) \to (2/\pi)^{1/2} \sin(kr - l\pi/2) \quad \text{as } r \to \infty,\ k > 0, \tag{1.2b}$$

$$u_i(r,k) \to (2/\pi)^{1/2} \sin(kr - (\alpha/k)\log(2kr) - l\pi/2 + \delta_i^l) \quad \text{as } r \to \infty,\ i = 1,2,\ k > 0$$

and $\delta_1^l = \arg[\Gamma(l + 1 + i\alpha/k)]$, $\alpha = e_1 e_2 m$. The position of the continuous spectrum of H_i for all i and l is $[0,\infty)$ and is absolutely continuous.

We denote the transform

$$f_i(k) = (F_i f)(k) = \text{l.i.m.} \int_0^\infty u_i(r,k) f(r)\,dr, \qquad f \in \mathcal{L}^2(0,\infty) \tag{1.3}$$

$$(F_i^* g)(r) = \text{l.i.m.} \int_0^\infty u_i(r,k) g(k)\,dk, \qquad g \in \mathcal{L}^2(0,\infty).$$

Note that F_i is partially isometric and

$$F_i^* F_i = P_i, \qquad F_i F_i^* = P_i', \tag{1.4}$$

where P_i is the projection from $\mathcal{L}^2(0,\infty)$ onto the continuous spectrum (which is absolutely continuous) subspace of H_i and P_i' is the projection from $\mathcal{L}^2(0,\infty)$ onto $F_i(\mathcal{L}^2(0,\infty))$. Formulas relating one-dimensional and three-dimensional quantities will be found in Appendix A.

We will also consider various generalized wave operators acting in $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^3)$. We let

$$W_\pm (H_j, H_i) = \operatorname*{s-lim}_{t\to\pm\infty} e^{iH_j t} e^{-iH_i t} P_i, \quad i,j = 1,2, \quad (1.5a)$$

where H_i, H_j are self-adjoint operators and P_i is the projection operator on the subspace of absolute continuity of H_i. We also will use the modified wave operators introduced by Dollard[2] which we denote by

$$W_\pm (H_1, H_D) = \operatorname*{s-lim}_{t\to\pm\infty} e^{iH_i t} e^{-iH_D(t)}, \quad i = 1,2, \quad (1.5b)$$

where $e^{-iH_D(t)}$ is unitary (but not a one-parameter group) and

$$H_D(t) = H_0 t + \epsilon(t) 1/2 e_1 e_2 (H_0)^{-1} \log(-4|t|H_0), \quad (1.6)$$

where $\epsilon(t) = \pm 1$ for $t \gtrless 0$. When the wave operators $W_\pm (H_j, H_f)$, $j = 1,2$, $f = 1,2,D$ exist and $R(W_\pm)$ are the absolutely continuous (continuous) spectrum subspace of H_j, then $W_\pm (H_j, H_f)$ will be said to be absolutely continuous (continuous) complete, abbreviated acc (cc). When considering these operators acting on $\mathcal{H}^l = \sum_{l=0}^\infty \oplus \mathcal{H}^l$ they will be denoted by $W_\pm = \sum_{l=0}^\infty \oplus W_\pm^l$. This notation is justified as a result of Kuroda[8] which shows that $W_\pm$ exist if and only if $W_\pm^l$ exist for each l and when $W_\pm$ exist, $W_\pm = \sum_{l=0}^\infty \oplus W_\pm^l$.

2. RESULTS

In part A of this section we give results on the existence and completeness of wave operators deduced from the time-dependent wave operator approach. In part B we give time-independent results and the relation between the time-dependent and time-independent results.

A. Time-Dependent Scattering

Contained in Dollard's work is the following:

Theorem 2.1[2]: In $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^3)$ we have that

(a) The wave operators $W_\pm(H_1, H_D)$ exist and are acc and cc.

(b) For $V \in \mathcal{L}^2(\mathbb{R}^3)$ and real, $W_\pm(H_2, H_D)$ exist.

We give the following extension of Theorem 2.1.

Theorem 2.2: Let $V = V_2 + V'$ with $V_2 \in \mathcal{L}^2(\mathbb{R}^3)$ and $V' \in \mathcal{L}^\infty(\mathbb{R}^3) \cap \mathcal{L}^p(\mathbb{R}^3)$ ($p < 3$). Then the wave operators $W_\pm(H_2, H_D)$ exist.

With H_0, H_1, H_2 defined as direct sums over l and H_1^l, H_2^l defined as operator sums with $\mathfrak{D}(H_i^l) = \mathfrak{D}(H_0^l)$, $i = 1,2$, we state a previously obtained result as the following theorem.

Theorem 2.3[11]*:* If V is spherically summetric and satisfies the condition

$$\int_0^R r^2 |V(r)|^2 dr + \int_R^\infty (1+r)^\delta |V(r)|^i dr < \infty, \qquad i = 1,2, \quad (2.1)$$

for some R $(0 \le R < \infty)$ and some δ $(0 < \delta < 1)$, then the generalized wave operators $W_\pm(H_2, H_1)$ and $W_\pm(H_1, H_2)$ exist on $\mathcal{L}^2(\mathbb{R}^3)$ and are acc. Furthermore, if $V(r)$ also satisfies Eqs. (1a)–(1c), the above wave operators are cc.

Remark 2.1: In proving the first part of the above theorem we applied a theorem of Kuroda[12] for operators. Kuroda[13] has an analogous theorem for forms which he used to extend Green and Lanford's results[7] for the case of the wave operators $W_\pm(H_2, H_1)$ with $V_c = 0$. We have not been able to obtain the required estimates to apply this theorem which would allow a $r^{-2+\epsilon}$ $(\epsilon > 0)$ behavior for V at the origin. In Appendix B we give an argument which indicates but does not prove existence and absolutely continuous completeness in this case.

Remark 2.2: With the condition of Eqs. (1a)–(1c) imposed on V the continuous spectrum of H_2 is absolutely continuous from Kodaira's theory of eigenfunction expansions.[4] Precisely to what extent the continuity assumption of V can be relaxed and still have cc has not been investigated.

Combining a modified version of the chain rule for generalized wave operators and Theorem 2.1, we have the following:

Theorem 2.4:

(a) Let V satisfy the conditions of (2.1). Then the wave operators $W_\pm(H_2, H_D)$ exist and are acc.

(b) If V satisfies both (2.1) and (1a)–(1c) then $W_\pm(H_2, H_D)$ exist and are cc.

(c) In both (a) and (b) the scattering operator $S = W_+^* W_-$ is unitary.

B. Time-Independent Scattering and the Relation to Time-Dependent Scattering

We give these results in one-dimensional form and suppress the l dependence. The three-dimensional results are obtained by taking direct sums. Dollard's results on cc for the wave operators $W_\pm(H_1, H_D)$ can then be stated as Theorem 2.5.

Theorem 2.5[2]*:* Defining

$$(U_\pm(H_1, H_D)f)(r) = \text{l.i.m.} \int u_1(r,k) e^{\mp i\delta_1(k)} (F_0 f)(k) dk$$

(l.i.m. with respect to the k integral means norm limit over a sequence $[\alpha, \beta]$ of k intervals, $\alpha > 0$, $\beta < \infty$, with $\alpha \to 0$ and $\beta \to \infty$) or

$$U_\pm(H_1, H_D)f = F_1^* e^{\mp i\delta_1} F_0 f, \quad (2.2)$$

then $U_\pm(H_1, H_D) = W_\pm(H_1, H_D)$.

We are now in a position to establish the time-independent theory for asymptotically Coulomb potentials and the relationship to the time-dependent theory. We have the following theorem.

Theorem 2.6:

(a) Let V satisfy (1.1a)–(1.1c) and define $S' = F_0^* e^{2i\delta_2} F_0$; then S' is unitary.

(b) Let V satisfy (1a)–(1c) and (2.1). Define $U_\pm(H_2, H_1) = F_2^* e^{\pm i(\delta_1 - \delta_2)} F_1$ on $\mathcal{L}^2(0,\infty)$; then $U_\pm(H_2, H_1) = W_\pm(H_2, H_1)$.

(c) With V as in part b we have $W_\pm(H_2, H_D) = F_2^* e^{\mp i\delta_2} F_0$. Define $S = W_+^*(H_2, H_D) W_-(H_2, H_D)$; Then $S = S'$.

Remark 2.3: Knowing the eigenfunction expansion for H_2, such as having δ_2, we can always define a unitary operator by $F_0^* e^{2i\delta_2} F_0$. The important point is that the time-independent definition agrees with the operator S defined from the time-dependent theory.

Remark 2.4: We expect the equality $W_\pm(H_2,H_1) = U_\pm(H_2,H_1)$ to be valid under the less restrictive conditions on V of (1a)–(1c) or more generally when an eigenfunction expansion is valid for H_2 and the continuous spectrum is absolutely continuous.

3. PROOF OF THE THEOREMS

For the proof of Theorems 2.1 and 2.5 see Dollard.[2] For the proof of Theorem 2.3 see O'Carroll.[11]

Proof of Theorem 2.2: We apply Theorem 3.7, p. 533 of Kato.[9] For a fundamental subset in $L^2(R^3)$ we take the space of functions $\mathcal{C} \subset S(R^3)$ used by Dollard,[2] where $f(x) \in \mathcal{C}$ if $f_0(q)$ vanishes in a neighborhood of $q = 0$. For $h \in \mathcal{C}$ we show that

$$\|(V_2 + V')e^{-iH_D(t)}h\|_2 \le \|V_2 e^{-iH_D(t)}h\|_2 + \|V' e^{-iH_D(t)}h\|_2 \quad (3.1)$$

is integrable on $[t_0, \infty]$ where $t_0 > 1$. We use Lemma 2 of Dollard[2]: Let $h \in \mathcal{C}$. Then, in the system of units used in the introduction for $|t| \ge t_0$, we have

$$(e^{-iH_D(t)}h)(x) = (1/2\mathrm{i}t)^{3/2}\phi_c(x)h_0(x/2t) + (\tfrac{1}{4}\pi i t)^{3/2} e^{ix^2/4t} R_h(x,t), \quad (3.2)$$

where

$$\phi_c(x) = \exp(ix^2/4t)\exp\{-i\epsilon(t)e_1e_2\log(x^2/|t|)t/2x\}$$

and for any integer n there exists a number $\mu \ge 0$ and a constant K, depending on h, such that

$$R_h(x,t) \le K(\log|t|)^\mu/|t|^{1/2}[1 + (x/t)^2]^n. \quad (3.3)$$

Returning to (3.1) we find

$$\|V_2 e^{-iH_D(t)}h\|_2 \le \|V_2\|_2 \|e^{-iH_D(t)}h\|_\infty \le \|V_2\|_2\{(2t)^{-3/2}\|h\|_1 + K(\log|t|)^\mu t^{-2}\} \quad (3.4)$$

which is integrable on $[t_0, \infty)$. The second term on the right side of (3.1) is found to satisfy

$$\begin{aligned}
&\|V' e^{-iH_D(t)}h\|_2 \\
&\le \|V'\phi_c(x)(2t)^{-3/2}h_0(x/2t)\|_2 + \||V'|(2t)^{-3/2}|R_h|\|_2 \\
&= \|V'^2(2t)^{-3}|R_h|^2\|_1^{1/2} + I' \\
&\le \|V'^2\|_p^{1/2} t^{-3/2}(\log|t|)^\mu t^{-1/2} t^{3/2p'} I^{1/2p'} + I' \\
&\le K''\{(\log|t|)^\mu/t^{2-3/2p'}\}\|V'\|_{2p} + I', \qquad (3.5)
\end{aligned}$$

where

$$p^{-1} + p'^{-1} = 1, \qquad I = \int_0^\infty y^2(1+y^2)^{-2p'n}dy$$

and

$$I' = \|V'\phi_c(x)(2t)^{-3/2}h_0(x/2t)\|_2.$$

We have used the Hölder inequality in going from the first to the second line of (3.5). For $p < 3/2$, $p' > 1$ so that the first term of (3.5) is integrable on $[t_0, \infty)$. The second term can be taken care of by the use of a formula due to B Simon. Using the fact that

$$\|(2t)^{-3/2}h_0(x/2t)\|_2 = \|h_0\|_2 = \|h\|_2,$$
$$\|(2t)^{-3/2}h_0(x/2t)\|_\infty \le (2t)^{-3/2}\|h\|_1,$$

and the Riesz–Thorin[14] convexity theorem we obtain Simon's formula:

$$\|(2t)^{-3/2}h_0(x/2t)\|_{p'} \leqq K t^{-3(1-2/p')/2}\|h\|_p$$

for $p^{-1} + p'^{-1} = 1$, $1 \le p \le 2$. Thus with $s^{-1} + s'^{-1} = 1$ and using the Hölder inequality we have

$$\begin{aligned}
&\|V'\phi_c(x)(2t)^{-3/2}h_0(x/2t)\|_2 \\
&\le \|V'^2\|_s^{1/2}\|(2t)^{-3/2}h_0(x/2t)\|_{2s} \\
&\le K'\|V'\|_{2s'} t^{-3/2s'}. \qquad (3.6)
\end{aligned}$$

Equation (3.6) is integrable for $s' < 3/2$ as V' is assumed to have finite p norm for some $p < 3$.

Proof of Theorem 2.4: (a) From Theorem 2.1a we have that $W_\pm(H_1,H_D)$ are acc and cc. From Theorem 2.3 we have that $W_\pm(H_2,H_1)$ are acc. We prove the chain rule (valid although $H_D(t)$ is not a one-parameter unitary group):

$$W_\pm(H_2,H_1)W_\pm(H_1,H_D) = W_\pm(H_2,H_D). \quad (3.7)$$

Let $P_1(P_2)$ be the projection on the absolutely continuous spectrum subspace of $H_1(H_2)$. As $W_\pm(H_1,H_D)\mathcal{H} = P_1\mathcal{H}$ and $W_\pm(H_2,H_1)$ are acc then $W_\pm(H_2,H_D)\mathcal{H} = P_2\mathcal{H}$, i.e., $W_\pm(H_2,H_D)$ are acc. We establish (3.7) for $t \to +\infty$. We have $W_+(H_2,H_1)W_+(H_1,H_D) = $ s–lim $e^{itH_2}P_1e^{-iH_D(t)}$ as $t \to \infty$. For $v \in L^2(R^3)$, we find

$$\begin{aligned}
&\|W_+(H_2,H_D)v - W_+(H_2,H_1)W_+(H_1,H_D)v\| \\
&= \|\lim_{t\to\infty}(e^{iH_2t}Ie^{-iH_D(t)}v - e^{iH_2t}P_1e^{-iH_D(t)}v)\| \\
&= \lim_{t\to\infty}\|(I-P_1)e^{-iH_D(t)}v\| \\
&= \lim_{t\to\infty}\|e^{iH_1t}(I-P_1)e^{-iH_D(t)}v\| \\
&= \|(I-P_1)W_+(H_1,H_D)v\|. \qquad (3.8)
\end{aligned}$$

Equation (3.8) is zero since $W_+(H_1,H_D)$ is acc which implies $W_+(H_1,H_D) = P_1W_+(H_1,H_D)$.

(b) From Kodaira's theory of eigenfunction expansions,[4] the absolutely continuous spectrum subspace of H_2 corresponds with the continuous spectrum subspace that $W_\pm(H_2,H_D)$ are cc.

(c) In both (a) and (b), $W_\pm\mathcal{H} = P_2\mathcal{H}$ so that $S = W_+^* W_-$ is unitary.

Proof of Theorem 2.6: We first give two lemmas which will be used in the proof of part b.

Lemma 3.1. Let $u(r,k)$ be the solution of

$$[-d^2/dr^2 + l(l+1)/r^2 + V(r)]u = k^2u$$

used in the eigenfunction expansions of Sec. 1 with V satisfying the conditions of (1a)–(1c) such that

$$u(0,k) = 0$$

and

$$u(r,k) \simeq (2/\pi)^{1/2}\sin[kr - l\pi/2 + (\alpha/k)\ln 2kr + \delta(k)], \qquad r \to \infty.$$

Then $u(r,k)$ is continuous in $((r,k): r \geq 0, k > 0)$ and is bounded in the set $D' = (r,k): 0 \leq r \leq r_2$, $0 < k_1 \leq k \leq k_2)$ with the bound

$$|u(r,k)| \leq C r^{l+1} e^{r^{\epsilon}/\epsilon}, \tag{3.9}$$

where $V(r) \sim O(1/r^{2-\epsilon})$ as $r \to 0$.

Proof of Lemma 3.1: Following Sec. 22.25 of Titchmarsh[15] we solve the integral equation

$$\begin{aligned} y(r,k) &= k^{-l-1/2} r^{1/2} J_{l+1/2}(kr) - (\pi/2) \\ &\times \int_0^\infty [J_{l+1/2}(kr) Y_{l+1/2}(ks) - J_{l+1/2}(kr) \\ &\times Y_{l+1/2}(ks)] r^{1/2} s^{1/2} V'(s) y(s,k) ds \end{aligned} \tag{3.10}$$

by iteration with $V'(r) = \alpha/r + V(r)$, and $y(r,k)$ is the solution which behaves as r^{l+1} as $r \to 0$. We easily find $y(r,k) \leq C' r^{l+1} e^{r^{\epsilon}/\epsilon}$ in D'. From Theorem 5.3 of Kodaira[4]

$$u(r,k) = (2/\pi)^{1/2} \cdot |k/A(k)| \, y(r,k),$$

where $A(k)$ is continuous in $[k_1, k_2]$ and nonzero so the theorem follows.

Lemma 3.2: Let V satisfy (1a), (1b), and (2.1) and have compact support.

Let

$$h(r, k^2 + i\sigma) = (R_{1, k^2 + i\sigma} V u_2)(r,k),$$

where the resolvent operator $R_{1,k^2+i\sigma} = (H_1 - k^2 - i\sigma)^{-1}$ is an integral operator in $L^2(0,\infty)$ with kernel $G_1(r, r', k^2 + i\sigma)$ $(\sigma \neq 0)$. Then

$$\lim_{\sigma \to 0+} h(r, k^2 + i\sigma) = (R_{1,k^2} V u_2)(r,k)$$

uniformly in any $D = ((r,k): 0 < r_1 \leq r \leq r_2$, $0 < k_1 \leq k \leq k_2)$, where R_{1,k^2} is the integral operator with kernel $G_1(r, r', k^2)$.

Proof of Lemma 3.2: From Theorem 20.21 of Stone[10] the resolvent operator $(H_1 - \lambda)^{-1}$, $\mathrm{Im}\lambda \neq 0$ is an integral operator of the Carleman type, denoted by $G_1(r,s,\lambda)$, given by

$$G_1(r,s,\lambda) = \begin{cases} u_1(r,\tau) w_1(s,\tau)/W(w_1,u_1), & 0 < r < s \\ w_1(r,\tau) u_1(s,\tau)/W(w_1,u_1), & 0 < s < r \end{cases}, \tag{3.11}$$

where $u_1(r,\tau)$ and $w_1(r,\tau)$ are independent solutions of $(L_1 - \lambda) y = 0$, $\tau = (\lambda)^{1/2}$, and $W(w_1, u_1)$ is the Wronskian. In Messiah's[1] notation, two linearly independent solutions are

$$\begin{aligned} F_l(\alpha/\tau, \tau r) &= c_l e^{i\tau r} (\tau r)^{l+1} \\ &\quad \times F(l + 1 + i(\alpha/\tau) | 2l + 2 | -2i\tau r), \\ G_l(\alpha/\tau, \tau r) &= v_1(r,\tau) \\ &= c_l e^{i\pi r} (\tau r)^{l+1} G(l + 1 + i\alpha/\tau | 2l + 2 | -2i\tau r) \end{aligned} \tag{3.12}$$

with $\mathrm{Re}\tau > 0$. F and G are confluent hypergeometric functions and

$$c_l(1/\tau) = [2^l/2l + 1)!] \exp[-i\pi\alpha/\tau] |\Gamma[l + 1 = (i\alpha/\tau)|.$$

Since $F(\beta|\gamma|\rho)$ is entire in β and ρ, and $G(\beta|\gamma|\rho)$ is entire in β and analytic in ρ except for $-\infty < \rho \leq 0$ (see Lebedev,[16] pages 261-65), F_l and G_l are C^∞ in the three variables (r, k, σ) where $r \in [0,\infty)$ for F_l $r \in (0,\infty)$ for G_l and $k \in (0,\infty)$, where we set $\lambda = k^2 + i\sigma$, $\mathrm{Re}\tau > 0$. For $\sigma > 0$ we have

$$\begin{aligned} w_1(r,\tau) &= u^+(\alpha/\tau, \tau r) = [G_l(\alpha/\tau, \tau r) + i F_l(\alpha/\tau, \tau r)] \\ &\qquad \times [\arg\Gamma(l + 1 + (i\alpha/\tau))] \\ u_1(r,\tau) &= (2/\pi)^{1/2} F_l(\alpha/\tau, \tau r); W^{-1}(w_1, u_1) \\ &= (\pi/2)^{1/2} (1/\tau) [\arg\Gamma(l + 1 + i\alpha/\tau)]. \end{aligned} \tag{3.13}$$

The above considerations show that

$$\lim_{\sigma \to 0^+} u_1[r, (k^2 + i\sigma)^{1/2}] = u_1(r,k)$$

uniformly in any domain $0 < k_1 \leq k \leq k_2$, $0 \leq r \leq r_2$ and that

$$\lim_{\sigma \to 0^+} w_1(r, (k^2 + i\sigma)^{1/2}) = w_1(r,k)$$

uniformly in any domain $0 < k_1 \leq k \leq k_2$, $0 < r_1 \leq r \leq r_2$. Thus

$$\lim_{\sigma \to 0^+} G_1(r, s, k^2 + i\sigma) = G_1(r, s, k^2)$$

uniformly in any domain $0 < r_1 \leq r \leq r_2$, $0 \leq s \leq s_2$, $0 < k_1 \leq k \leq k_2$. Let $\mathrm{supp} V \subset [0, s_2]$. We have

$$\int_0^\infty |V(s) u_2(s,k)| ds \leq C' \int_0^{s_2} s^2 V^2(s) ds = C'' < \infty \tag{3.14}$$

so that

$$V(r) u_2(r,k) \in L^2(0,\infty) \quad \text{and} \quad V(r) u_2(r,k) \in L^1(0,\infty)$$

with a bound C'' which is uniform in $0 < k_1 \leq k \leq k_2$. We find

$$\begin{aligned} \sup_D & |h(r, (k^2 + i\sigma)^{1/2}) - h(r,k)| \\ &\leq \sup_D \int_0^{S_2} |G_1(r, s, k^2 + i\sigma) \\ &\quad - G_1(r, s, k^2)| |V(s) u_2(r,k)| ds \\ &\leq C'' \sup_D |G_1(r, s, k^2 + i\sigma) - G_1(r, s, k^2)|, \end{aligned}$$

which suffices to establish the lemma.

Proof of Theorem 2.6(a): This is obvious.

Proof of Theorem 2.6(b): As this proof is rather lengthy we give a brief outline. In Part I we prove the theorem for a continuous, cutoff V, denoted by V_n, with supp $V_n \subset [0, n+1]$. The equality $W_\pm(H_1 + V_n, H_1) = U_\pm(H_1 + V_n, H_1)$ is established following a method of Ikebe[17] and uses previous results on the existence and cc of the wave operators $W_\pm(H_2, H_1)$.[11] In Part II, following Kuroda[8], a limiting procedure is used which allows us to prove the result for V.

Part I: We will prove

$$W_\pm(H_1 + V_n, H_1) = F_2^* e^{\pm i(\delta_1 - \delta_2)} F_1 = U_\pm(H_1 + V_n, H_1)$$

(We suppress the n dependence of F_2^* and δ_2) for a continuous cut-off potential defined by

$$V_n(r) = \begin{cases} V(r), & 0 < r \leq n \\ [\mathrm{sgn}\, V(n)] \min(|V(r)|, |V(n)|(n + 1 - r)), \\ n \leq r \leq n+1 \quad 0, & n + 1 \leq r < \infty. \end{cases}$$

With this definition, $|V_n(r)| \le |V(r)|$. Throughout Part I we let $f \in C_0^\infty(0,\infty)'$ and g is such that $F_1 g \in C_0^\infty(0,\infty)'$, where $C_0^\infty(0,\infty)'$ is the set of all C^∞ functions with compact support contained in $(0,\infty)$. Furthermore we consider only the $t \to -\infty$ case, the proof for $t \to +\infty$ is similar. With f and g as above we follow Ikebe[17] (see Sec. 11) to establish that $U_\pm(H_1 + V_n, H_1)$ are well defined. We have

$$\begin{aligned}(U_- g)(r) &= \int_0^\infty u_2(r,k)\, e^{i(\delta_1-\delta_2)} (F_1 g)(k)\, dk \\ &= F_2^* e^{i(\delta_2-\delta_1)}\ F_1 g, \qquad (3.15)\\ (U_-^* f)(r) &= \int_0^\infty u_1(r,k)\, e^{i(\delta_1-\delta_2)}\ (F_2 f)(k)\, dk \\ &= F_1^* e^{i(\delta_1-\delta_2)}\ F_2 f,\end{aligned}$$

which can be extended to all $f, g \in L^2(0,\infty)$ by writing l.i.m. in front of the integrals. Note that

$$U_-(L^2(0,\infty)) \subset P_2(L^2(0,\infty)), \qquad (3.15'a)$$

$$U_-^*(L^2(0,\infty)) \subset P_1(L^2(0,\infty)). \qquad (3.15'b)$$

We also have that

$$U_-^* e^{iH_2 t} = e^{iH_1 t}\ U_-^*, \qquad U_-^* H_2 \subset H_1 U_-^*. \qquad (3.16)$$

We will now derive

$$U_-^*(H_1 + V_n, H_1)\ W_-(H_1 + V_n, H_1) = P_1.$$

Starting from

$$U_-^* W_- = \operatorname*{s-lim}_{t\to-\infty} U_-^* e^{i(H_1+V_n)t} e^{-tH_1} P_1 \qquad (3.17)$$

and, using the existence of the strong limit, the fact that $u_2 V_n \in L^2(0,\infty)$ and that $P_1 F_1 = F_1$, we arrive at

$$\begin{aligned}(U_-^* W_- f, P_1 g) = (U_-^* f, P_1 g) &+ \lim_{\delta\to 0^+} \int_0^\infty (f, R_{1,k^2+i\sigma} V_n u_2) \\ &\times (F_1 g)(k)\, e^{i(\delta_2-\delta_1)} dk. \qquad (3.18)\end{aligned}$$

By Lemma 3.2 we can pass the limit inside the integral and inner product of (3.18) obtaining

$$\begin{aligned}(U_-^* W_- f, g) &= (U_-^* W_- f, P_1 g) \\ &= \int_0^\infty e^{-i(\delta_1-\delta_2)} (F_1 g)(k) \int_0^\infty \bar{f}(r)\ \times \phi(r,k)\, dr\, dk \qquad (3.19)\end{aligned}$$

with

$$\begin{aligned}\phi(r,k) = u_2(r,k) &+ \int_0^{n+1} G_1(r,r',k^2)\, V_n(r') \\ &\times u_2(r',k)\, dr', \qquad k \ne 0,\end{aligned}$$

where $G_1(r,r',k^2)$ is given by (3.11). For fixed k $(k \ne 0)$ we see that $G_1(r,r',k^2)$ is a bounded continuous function in a neighborhood of the r, r' origin.

Thus $\phi(r,k)$ is continuous in $0 \le r < \infty$ and bounded at the origin. Furthermore, $(L_1 - k^2)\,\phi(r,k) = 0$ so $\phi(r,k) = c_k u_1(r,k)$. Using the explicit form of $G_1(r,r',k^2)$ [see (3.11) and (3.13)], we have

$$\phi(r,k) = (u_2 + K w_1)(r,k) \qquad \text{for } r \ge n+1.$$

Thus, comparing the e^{ikr} part of the asymptotic form of $u_2(r,k)$ and $u_1(r,k)$ for $r \to \infty$, we find $c_k = e^{i(\delta_2-\delta_1)}$. Equation (3.19) becomes $(U_-^* W_- f, P_1 g = \int_0^\infty (F_1 g)(k) \int_0^\infty \bar{f}(r)\, u_1(r,k)\, dr\, dk = (P_1 f, g)$. By (3.15b),

$$P_1 U_-^* = P_1 \quad \text{so} \quad (U_-^* W_- f, g) = (P_1 f, g)$$

for arbitrary f and g dense in $L^2(0,\infty)$. Thus we have the $L^2(0,\infty)$ relation

$$U_-^* W_- = P_1. \qquad (3.20)$$

Using (3.20), we see that it is easy to show $U_- = W_-$. From Theorem 2.3 it follows that W_- is acc, so that $W_- W_-^* = P_2$. Thus

$$U_-^* W_- W_-^* = P_1 W_-^* = U_-^* P_2.$$

Taking adjoints we have $P_2 U_- = W_- P_1$. Using (3.15a) we obtain $U_- = W_-$. We have established on $L^2(0,\infty)$ that

$$W_\pm(H_1 + V_n, H_1) = F_2^{n*}\ e^{\pm i(\delta_1-\delta_2^n)} F_1. \qquad (3.21)$$

In (3.21) we have resurrected the n dependence of F_2^* and δ_2. In what follows F_2 and δ_2 will be associated with $H_2 = H_1 + V$.

Part II: We will follow a method of Kuroda[8] which allows passage to the limit $n \to \infty$ in (3.21), obtaining $U_-(H_2, H_1) = W_-(H_2, H_1)$. We will prove the following:

$$\operatorname*{s-lim}_{n\to\infty} W_\pm(H_1 + V_n, H_1) = W_\pm(H_2, H_1), \qquad (3.22)$$

$$\lim_{n\to\infty} \delta_2^n(k) = \delta_2(k) \quad [\text{uniformly in } (0,\infty)], \qquad (3.23)$$

$$\operatorname*{s-lim}_{n\to\infty} F_2^* e^{\pm i(\delta_1-\delta_2^n)} F_1 = F_2^*\ e^{\pm i(\delta_1-\delta_2)} F_1. \qquad (3.24)$$

From (3.21), (3.22), and (3.24) part b of the theorem follows.

Proof of (3.22): By Theorem 2 of Kuroda[12] it is sufficient to show that $|V - V_n|^{1/2} (H_1 - \lambda)^{-1}$ is Hilbert-Schmidt and

$$\lim_{n\to\infty} \|\, |V - V_n|^{1/2} (H_1 - \lambda)^{-1} \|_{H\text{-}S} = 0$$

for some λ (which we take to be real) in the resolvent set of H_1. Since $|(V - V_n)(r)| \le 2|V(r)|$ and

$$\|\, |V|^{1/2} (H_1 - \lambda)^{-1} \|_{H\text{-}S} < \infty,$$

we have

$$\begin{aligned}&\|\, |V - V_n|^{1/2} (H_1 - \lambda)^{-1} \|_{H\text{-}S}^2 \\ &\qquad \le 2 \|\, |V|^{1/2} (H_1 - \lambda) \|_{H\text{-}S}^2 < \infty.\end{aligned}$$

The Lebesgue dominated convergence theorem gives

$$\begin{aligned}&\lim_{n\to\infty} \|\, |V - V_n|^{1/2}\ (H_1 - \lambda)^{-11} \|_{H\text{-}S}^2 \\ &\qquad = 0 \text{ so that } \operatorname*{s-lim}_{n\to\infty} W_\pm(H_1 + V_n, H_1) = W_\pm(H_2, H_1).\end{aligned}$$

Proof of (3.23): Let $u_2^n(r,k)$ satisfy $(L_1 + V_n - k^2)\, u_2^n = 0$ with the boundary and asymptotic conditions

$$\lim_{r\to 0} u_2^n(r,k) = 0, \quad u_2(r,k) \sim (2/\pi)^{1/2} \sin[kr - (l\pi/2) + (\alpha/k)\ln 2(kr) - \delta_2^n(k)]$$

as $r \to \infty$. Let $\tilde{u}_2^n(r,k)$ also satisfy

$$(L_1 + V_n - k^2)\tilde{u}_2^n = 0 \quad \text{and} \quad \tilde{u}_2^n(r,k) = u_2(r,k) \quad \text{for } 0 \leqslant r \leqslant n.$$

We find for $r \geqslant n+1$

$$(\tilde{u}_2^n - u_2)(r,k) = W(u_1,v_1)^{-1}\left(\int_n^{n+1} K_1(r,s,k)[V(s)u_2(s,k) - V_n(s)u_2^n(s,k)]ds + \int_{n+1}^r K_1(r,s,k)V(s)u_2(s,k)ds\right) \quad (3.25)$$

with $K_1(r,s,k) = u_1(r,k)v_1(s,k) - u_1(s,k)v_1(r,k)$.

Kuroda[8] derives a formula analogous to (3.25) for the case $\alpha = 0$ which is correct only for $l = 0$. However, using the appropriate kernel for $l \neq 0$, his argument goes through. Since $V_n(r) = V(r)$ for $0 \leqslant r \leqslant n$, we readily obtain

$$\tilde{u}_2^n(r,k) = c_{nk} u_2^n(r,k), \quad (3.26)$$

where the c_{nk} can be chosen positive.

Starting from (3.25) we derive

$$\lim_{n\to\infty} (\tilde{u}_2^n - u_2)(r,k) = 0 \text{ uniformly in } 0 \leqslant r < \infty \text{ and } 0 < k_1 \leqslant k \leqslant k_2.$$

Hence, by virtue of the asymptotic forms of $\tilde{u}_n^2$ and u_2, and the periodic property of the sine function, we get

$$\lim_{n\to\infty} \{[\sin[kr - \tfrac{1}{2}l\pi + (\alpha/k)\ln 2kr - \delta_2] - c_{nk}\sin[kr - \tfrac{1}{2}l\pi + (\alpha/k)\ln 2kr - \delta_2^n]\} = 0$$

uniformly for all r and $0 < k_1 \leqslant k \leqslant k_2$. This readily implies

$$\lim_{n\to\infty} \delta_2^n(k) = \delta(k), \quad \lim_{n\to\infty} c_{nk} = 1 \quad (3.27)$$

uniformly for $0 < k_1 \leqslant k \leqslant k_2$. Then (3.26), (3.27) also show that

$$\lim_{n\to\infty} u_2^n(r,k) = u_2(r,k) \quad (3.28)$$

uniformly in r and in $0 < k_1 \leqslant k \leqslant k_2$.

Proof of (3.24): Kuroda[8] states that (3.24) can be proved (when $\alpha = 0$) using a standard argument. We have not been able to construct his argument, so we prove (3.24) directly. Let

$$g = F_1 f, \quad g \in C_0^\infty(0,\infty)'$$

and set (for the $t \to -\infty$ case)

$$h_n(r) = (F_2^{n*} e^{(\delta_2^n - \delta_1)} F_1 f)(r) = \int_0^\infty u_2^n(r,k)\, e^{(\delta_2^n - \delta_1)} g(k)\,dk.$$

From (3.27) and (3.28), we have

$$\lim_{n\to\infty} h_n(r) = h(r) = \int_0^\infty u_2(r,k)\, e^{i(\delta_2 - \delta_1)} g(k)\,dk$$

uniformly in $r \in [0,\infty)$. Equations (3.21) and (3.22) guarantee the existence of an h' such that

$$\lim_{n\to\infty} \|h_n - h'\| = 0.$$

Equation (3.24) is readily established by showing that $h = h'$.

Proof of Theorem 2.6(c): We first show that

$$W_\pm(H_2,H_D) = W_\pm(H_2,H_1)W_\pm(H_1,H_D) = F_2^* e^{\pm i(\delta_1 - \delta_2)} F_1 F_1^* e^{\mp i\delta_1} F_0 = F_2^* e^{\mp i\delta_2} F_0. \quad (3.29)$$

To show (3.29), it is sufficient to show that $P_1' = F_1F_1^*$ acts as the identity in $L^2(0,\infty)$. This follows since

$$(u, W_\pm^*(H_1,H_D)W_\pm(H_1,H_D)v) = (u,v) = (e^{\mp i\delta_1}F_0 u, e^{\mp i\delta_1}F_0 v) = (e^{\mp i\delta_1}F_0 u, F_1F_1^* e^{\mp i\delta_1}F_0 v)$$

and $e^{\mp i\delta_1}F_0$ is onto $L^2(0,\infty)$. To show that

$$S = W_+^*(H_2,H_D)W_-(H_2,H_D) = F_0^* e^{i\delta_2} F_2F_2^* e^{i\delta_2} F_0 = F_0^* e^{i2\delta_2} F_0, \quad (3.30)$$

it is sufficient to show that $F_2F_2^*$ acts as the identity in $L^2(0,\infty)$ in (3.30). This follows from (3.29) and the fact that $e^{\mp i\delta_2}F_0$ is onto $L^2(0,\infty)$ since

$$(u, W_\pm^*(H_2,H_D)W_\pm(H_2,H_D)v) = (u,v) = (e^{\mp i\delta_2}F_0 u, e^{\mp i\delta_2}F_0 v) = (e^{\mp i\delta_2}F_0 u, F_2F_2^* e^{\mp i\delta_2}F_0 v).$$

APPENDIX A

We give formal expressions in three- and one-dimensional form corresponding to those used in the text. A generalized Fourier transform in three dimensions we write as

$$f_i(q) = \int \overline{\Phi}_i(x,q) f(x)\,dx, \quad i = 0,1,2. \quad (A1)$$

Equation (A1) gives a mapping from $\mathcal{L}^2(\mathbb{R}^3)$ onto the continuous spectrum subspace of H_i. In addition to the kernel $\Phi_i(x,q) \equiv \Phi_i^-(x,q)$, we also have the related kernel $\overline{\Phi}_i(x,-q) \equiv \Phi_i^+(x,q)$. These kernels have the expansions

$$\Phi_i^\pm(x,q) = (kr)^{-1} \times \sum_{\substack{l=0\\|m|<l}}^\infty i^l\, e^{\mp i\delta_i^l(k)}\, u_i^l(r,k)\,\overline{Y}_{lm}(\hat{q})\,Y_{lm}(\hat{x}). \quad (A2)$$

The radial functions $u_i(r,k)$ have a phase shifted e^{ikr} asymptotic part for $i = 1,2$; for $i = 0$, $\delta_0^l(k) = 0$ for all l, $\delta_1^l(k) = \arg\Gamma[l + 1 + (i\alpha/k)]$ with $\alpha = e_1e_2m$. $\Phi_0^\pm(x,q) = (2\pi)^{-3/2} e^{iq\cdot x}$, the kernel of the $\mathcal{L}^2(\mathbb{R}^3)$ Fourier transform, and

$$\Phi_1^-(x,q) = (2\pi)^{-3/2} e^{iq\cdot x} e^{-\pi\alpha/2k} \times \Gamma(1+i\alpha/k)\, F(-i\alpha/k|1|ikr - iq\cdot x), \quad (A3)$$

where F is the confluent hypergeometric function. The asymptotic form of $\Phi_1^-(x,q)$ is

$$\Phi_1^-(x,q) \sim \Phi_{1_{\rm inc}}(x,q) + \Phi_{1_{\rm sc}}(x,q), \qquad |r - \hat{q}\cdot x| \to \infty, \quad (A4a)$$

$$\Phi_{1_{\rm inc}}^-(x,q) = (2\pi)^{-3/2} \exp\{i[q\cdot x + \alpha \log(kr - q\cdot x)/k]\} \times [1 + (\alpha/ik^2)(kr - q\cdot x) + \cdots], \quad (A4b)$$

$$\begin{aligned}\Phi_{1_{\rm sc}}^-(x,q) &= (2\pi)^{-3/2}(-\alpha/k)(kr - q\cdot x)^{-1}\,\Gamma(1+i\alpha/k) \\ &\times (\Gamma(1-i\alpha/k))^{-1} \exp\{i[kr - \alpha \log(kr - q\cdot x)/k]\} \\ &= (2\pi)^{-3/2} r^{-1} \exp\{i[kr - \alpha \log(2kr)/k]\} f_{\rm ck},\end{aligned} \quad (A4c)$$

where $\cos\theta = \hat{q}\cdot\hat{x}$, $\hat{x} = \hat{q}'$, and

$$\begin{aligned}f_{\rm ck}(\theta) &= (-\alpha/k)[2k\sin^2(\theta/2)]^{-1}\exp\{-i(\alpha/k) \times \log[\sin^2(\theta/2)] + 2i\delta_1^0\} \\ &= (-\alpha/k)(2k)^{-1}[(1-\cos\theta)/2]^{-1-i\alpha/k} \times \exp(2i\delta_1^0)\end{aligned} \quad (A4d)$$

is identified as the Coulomb scattering amplitude.

As sesquilinear forms, the $\Phi_i^\pm(x,q)$ obey

$$\int \overline{\Phi}_i^\pm(x,q')\,\Phi_i^\pm(x,q)\,dx = \delta(q'-q) + k^{-2}\delta(k'-k) \sum_{\substack{l=0\\|m|<l}}^{\infty} Y_{lm}(\hat{q})\, Y_{lm}(\hat{q}'). \quad (A5)$$

The time-independent operators of Eq. (2.2) in three-dimensional form are

$$(U_\pm(H_1,H_D)f)(x) = \int \Phi_1^\pm(x,q)\left(\int \overline{\Phi}_0(x',q) f(x')\,dx'\right) dq, \quad (A6)$$

$$(U_\pm^*(H_1,H_D)f)(x) = \int \Phi_0(x,q)\left(\int \overline{\Phi}_1^\pm(x',q) f(x')\,dx'\right) dq. \quad (A7)$$

Also in (2.3) we have

$$(U_\pm(H_2,H_1)f)(x) = \int \Phi_2^\pm(x,q)\left(\int \overline{\Phi}_1^\pm(x',q) f(x')\,dx'\right) dq \quad (A8)$$

$$(U_\pm^*(H_2,H_1)f)(x) = \int \Phi_1^\pm(x,q)\left(\int \Phi_2^\pm(x',q) f(x')\,dx'\right) dq. \quad (A9)$$

Then with $U_\pm(H_2,H_D) = U_\pm(H_2,H_1)\,U_\pm(H_1,H_D)$, we have

$$(U_\pm(H_2,H_D)f)(x) = \int \Phi_2^\pm(x,q) f_0(q)\,dq. \quad (A10)$$

The S' operator of Theorem 1.6 as a sesquilinear form in three-dimensions is

$$\begin{aligned}(g,S'f) &= (U_+(H_2,H_D)g,\, U_-(H_2,H_D)f) \\ &= \int \overline{g}_0(q')\left(\int \overline{\Phi}_2^+(x,q')\,\Phi_2^-(x,q)\,dx\right) f_0(q)\,dq\,dq'.\end{aligned} \quad (A11)$$

Taking into account (A2) we have the one-dimensional form of (A6)–(A10) where we will suppress the l dependence:

$$\begin{aligned}(U_\pm(H_1,H_D)f)(r) &= \int_0^\infty e^{\mp i\delta_1(k)} u_1(r,k)(F_0 f)(k)\,dk \\ &\equiv (F_1^* e^{\mp i\delta_1} F_0 f)(r),\end{aligned} \quad (A6')$$

$$(U_\pm^*(H_1,H_D)f)(r) = (F_0^* e^{\pm i\delta_1} F_1 f)(r), \quad (A7')$$

$$(U_\pm(H_2,H_1)f)(r) = (F_2^* e^{\mp i(\delta_2-\delta_1)} F_1 f)(r), \quad (A8')$$

$$(U_\pm^*(H_2,H_1)f)(r) = (F_1^* e^{\pm i(\delta_2-\delta_1)} F_2 f)(r), \quad (A9')$$

$$(U_\pm(H_2,H_D)f)(r) = (F_2^* e^{\mp i\delta_2} F_0 f)(r). \quad (A10')$$

Now returning to (A11) we substitute the expansions of (A2) to obtain

$$(g,S'f) = \int \overline{g}_0(q')\,k^{-2}\,\delta(k'-k) \times \sum_{\substack{l=0\\|m|\le l}}^{\infty} [e^{2i\delta_2^l(k)}\, \overline{Y}_{lm}(\hat{q})\, Y_{lm}(\hat{q}')] f_0(q)\,dq\,dq'. \quad (A12)$$

Using (A5) and $\delta(E'-E) = (2k)^{-1}\delta(k'-k)$, we then have

$$(g,(S'-I)f) = \int \overline{g}_0(q')[-2\pi i\,\delta(E'-E) \times T(q',q;k=k')] f_0(q)\,dq\,dq', \quad (A13)$$

where the scattering amplitude $f_k(\hat{q}\cdot\hat{q}')$ is given by

$$f_k(\hat{q}\cdot\hat{q}') = -2\pi^2\,T(q',q;k=k') = (2ik)^{-1}\sum_{l=0}^{\infty}(2l+1) \times \{\exp[2i\delta_2^l(k)] - 1\}\,P_l(\hat{q}\cdot\hat{q}'). \quad (A14)$$

The scattering amplitude of (A14) makes up part of a kernel of a sesquilinear form and, as such, is well defined. However, from a physical point of view, we are interested in its properties as a function of k and $\hat{q}\cdot\hat{q}'$. Unfortunately we have not been able to rigorously show that (A14), in the case $V=0$, agrees with (A4d).

APPENDIX B

We give a nonrigorous argument that indicates that the wave operators $W_\pm(H_j,H_i)$ $(i,j=1,2)$ exist and are absolutely continuous complete under the assumption that V is spherically symmetric and satisfies

$$|V| \le K\,(r^{-2+\epsilon} + r^{-1-\epsilon}), \qquad \epsilon > 0.$$

We use a technique of Kuroda[12] which he successfully used for the case $V=0$. By Kuroda's theorem for forms to show absolutely continuous completeness and existence of the wave operators, it is sufficient to show that $|V|^{1/2}(H_1-\lambda)^{1/2}$ is Hilbert–Schmidt for some $\lambda < \inf(\text{spectrum of } H_1)$. Using the representation

$$\begin{aligned}&[(H_1-\lambda)^{1/2} f](r) \\ &= \lim_{n\to\infty}\left(\sum_{m=0}^{n} u_{lm}(r)(u_{lm},f)(\lambda_m-\lambda)^{-1/2} + \int_{n^{-1}}^{n} u_1(r,k) \times (F_1 f)(k)\,(k^2-\lambda)^{-1/2}\,dk\right), \\ &\qquad f \in C_0^\infty(0,\infty),\end{aligned}$$

we have $\||V|^{1/2}(H_1-\lambda)^{-1/2}\|_{H\text{-}S}^2 = S + I$, where

$$S = \sum_{n=0}^{\infty}(u_{ln}, |V|\,u_{ln})(\lambda_n-\lambda)^{-1},$$

$$I = \int_0^\infty \int_0^\infty |V(r)|\, u_1^2(r,k)\,(k^2-\lambda)^{-1}\,dk\,dr.$$

Choosing λ such that $\mathrm{supp}(\lambda_n - \lambda) < 1$, we find the bound

$$S \le \sum_n \left[(u_{ln}, r^{-2}u_{ln}) + (u_{ln}, r^{-1}u_{ln})\right]$$

$$\le \sum_n \left[(2l+1)^{-1}n^{-3} + n^{-2}\right] < \infty.$$

Using the more natural variables $p = kr$, we have

$$I = \int_0^\infty \int_0^\infty |V(p/k)|\, u_1^2(p,k)\,(k^2-\lambda)^{-1}k^{-1}\,dk\,dp,$$

where

$$|V(p/k)| \le K\,(k^{2-\epsilon}p^{-2+\epsilon} + k^{1+\epsilon}p^{-1-\epsilon}),$$

$$|V(p/k)|(k^2-\lambda)k^{-1} \le K' k^{\epsilon}(1+k)^{-1-2\epsilon}\,(1+p)^{1-2\epsilon}\,p^{\epsilon-2}.$$

Recalling that

$$u_1(p,k) \propto c_l e^{ip}p^{l+1}F(l+1+i\alpha/k\,|\,2l+2\,|-2ip),$$

where

$$c_l \propto e^{-\pi\alpha/2k}\,|\Gamma(l+1+i\alpha/k)| \propto [(\alpha/k)\,e^{-\pi\alpha/k}\sinh^{-1}(\alpha/k)]\,\Pi_{m=1}^{l}\,(m^2+\alpha^2/k^2),$$

we find that

$$c_l = \begin{cases} k^{-l-1/2}e^{-2\pi\alpha/k} \to 0, & k\to 0,\ \alpha/k > 0, \\ k^{-1/2}, & k \to 0,\ \alpha/k < 0, \\ 1, & k \to \infty. \end{cases}$$

For $p \to 0$, $k \to \infty$, $F(l+1+i\alpha/k\,|\,2l+2\,|-2ip) \to 1$; for $p\to\infty$, $k\to\infty$,

$$u_1(p,k) \sim (2/\pi)^{1/2}\sin[p - (\alpha/k)\log(2p) - \tfrac{1}{2}l\pi + \delta_1^l)]$$

$$\to (2/\pi)^{1/2}\sin(p - \tfrac{1}{2}l\pi,\ p \ge l(l+1) + (\alpha/k).$$

From Landau and Lifschitz[18] (see p. 150) we have for $k \to 0$,

$$u_1(r,0)\,k^{-1/2} \propto (2r)^{1/2}J_{2l+1}[(8r)^{1/2}].$$

These considerations indicate that

$$u_1(p,k) \le K\,(1+k)^{1/2}k^{-1/2}(1+p)^{-l-1}p^{l+1}.$$

If the above bound for u_1 holds, then $\mathcal{I} < \infty$.

1 A. Messiah, *Quantum Mechanics* (Wiley, New York, 1966), Vol. I (see Chap. XI).
2 J. Dollard, J. Math. Phys. **5**, 729 (1964).
3 J. Dollard, Commun. Math. Phys. **12**, 193 (1969).
4 K. Kodaira, Amer. J. Math. **71**, 921 (1949).
5 D. Mulherin and I. I. Zinnes, J. Math. Phys. **11**, 1402 (1970).
6 Amrein, Martin, and Misra, Helv. Phys. Acta **43**, 313 (1970).
7 T. A. Green and O. F. Lanford, J. Math. Phys. **1**, 139 (1960).
8 S. T. Kuroda, J. Math. Phys. **3**, 933 (1962).
9 T. Kato, *Pertubation Theory of Linear Operators* (Springer, New York, 1966).
10 M. H. Stone, *Linear Transformations in Hilbert Space* (Amer. Math. Soc., New York, 1966), Vol. XV (see Chap. X).
11 M. O'Carroll, J. Math. Phys. 13, 268 (1972).
12 S. T. Kuroda, J. Math. Soc. Japan **11**, 247 (1959).
13 S. T. Kuroda, J. Math. Soc. Japan **12**, 243 (1960).
14 R. E. Edwards, *Fourier Series* (Holt, Reinhart, and Winston, New York, 1967), Vol. II.
15 E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second Order Differential Equations* (Oxford U.P., New York, 1962), Parts I and II.
16 N. N. Lebedev, *Special Functions and Their Applications* (Prentice-Hall, New York, 1965).
17 T. Ikebe, Arch. Ratl. Mech. Anal. 5, 1 (1960).
18 L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York)

On Weyl and Lyra Manifolds*

D. K. Sen and J. R. Vanstone

Department of Mathematics, University of Toronto, Toronto, Canada

(Received 1 November 1971; Revised Manuscript Received 6 February 1972)

It is shown that Weyl's geometry and an apparently similar geometry suggested by Lyra are special cases of manifolds with more general connections. The difference between the two geometries and their relationship with Riemannian geometry are clarified by giving a global formulation of Lyra's geometry. Finally the outline of a field theory based on the latter geometry is given.

1. INTRODUCTION

Shortly after Einstein's general theory of relativity Weyl[1] suggested the first so-called unified field theory based on a generalization of Riemannian geometry. In retrospect, it would seem more appropriate to call Weyl's theory a geometrized theory of gravitation and electromagnetism (just as the general theory was a geometrized theory of gravitation only), rather than a unified field theory. It is not quite clear to what extent the two fields have been unified, even though they acquire (different) geometrical significances in the same geometry. The theory was never taken seriously because it was based on the concept of nonintegrability of length transfer, and, as pointed out by Einstein, this implies that spectral frequencies of atoms depend on their past histories and therefore have no absolute significance. Nevertheless, Weyl's geometry provides an interesting example of non-Riemannian connections, and recently Folland[2] has given a global formulation of Weyl manifolds thereby clarifying considerably many of Weyl's basic ideas.

In 1951 Lyra[3] suggested a modification of Riemannian geometry which bears a remarkable resemblance to Weyl's geometry. But in Lyra's geometry, unlike Weyl's, the connection is metric preserving as in Riemannian geometry; in other words, length transfers are integrable. Lyra also introduced the notion of a gauge and in the "normal" gauge the curvature scalar is identical to that of Weyl. It is thus possible[4] to construct a geometrized theory of gravitation and electromagnetism much along the lines of Weyl's "unified" field theory without, however, the inconvenience of nonintegrable length transfer.

$$I = \int_0^\infty \int_0^\infty |V(r)|\, u_1^2(r,k)\,(k^2-\lambda)^{-1}\,dk\,dr.$$

Choosing λ such that $\operatorname{supp}(\lambda_n - \lambda) < 1$, we find the bound

$$S \le \sum_n \left[(u_{ln}, r^{-2}u_{ln}) + (u_{ln}, r^{-1}u_{ln})\right]$$
$$\le \sum_n \left[(2l+1)^{-1}n^{-3} + n^{-2}\right] < \infty.$$

Using the more natural variables $p = kr$, we have

$$I = \int_0^\infty \int_0^\infty |V(p/k)|\, u_1^2(p,k)\,(k^2-\lambda)^{-1}k^{-1}\,dk\,dp,$$

where

$$|V(p/k)| \le K\,(k^{2-\epsilon}p^{-2+\epsilon} + k^{1+\epsilon}p^{-1-\epsilon}),$$
$$|V(p/k)|(k^2-\lambda)k^{-1}$$
$$\le K' k^{\epsilon}(1+k)^{-1-2\epsilon}\,(1+p)^{1-2\epsilon}\,p^{\epsilon-2}.$$

Recalling that

$$u_1(p,k) \propto c_l e^{ip} p^{l+1} F(l+1+i\alpha/k \,|\, 2l+2 \,|\, -2ip),$$

where

$$c_l \propto e^{-\pi\alpha/2k}\,|\Gamma(l+1+i\alpha/k)|$$
$$\propto [(\alpha/k)\,e^{-\pi\alpha/k}\sinh^{-1}(\alpha/k)]\,\Pi_{m=1}^{l}\,(m^2+\alpha^2/k^2),$$

we find that

$$c_l = \begin{cases} k^{-l-1/2}e^{-2\pi\alpha/k} \to 0, & k\to 0,\ \alpha/k > 0,\\ k^{-1/2}, & k \to 0,\ \alpha/k < 0,\\ 1, & k\to\infty. \end{cases}$$

For $p \to 0$, $k \to \infty$, $F(l+1+i\alpha/k \,|\, 2l+2 \,|\, -2ip) \to 1$; for $p\to\infty$, $k\to\infty$,

$$u_1(p,k) \sim (2/\pi)^{1/2}\sin[p - (\alpha/k)\log(2p) - \tfrac{1}{2}l\pi + \delta_1^l)]$$
$$\to (2/\pi)^{1/2}\sin(p - \tfrac{1}{2}l\pi,\ p \ge l(l+1) + (\alpha/k).$$

From Landau and Lifschitz[18] (see p. 150) we have for $k \to 0$,

$$u_1(r,0)\,k^{-1/2} \propto (2r)^{1/2} J_{2l+1}[(8r)^{1/2}].$$

These considerations indicate that

$$u_1(p,k) \le K\,(1+k)^{1/2}k^{-1/2}(1+p)^{-l-1}p^{l+1}.$$

If the above bound for u_1 holds, then $\mathscr{I} < \infty$.

1 A. Messiah, *Quantum Mechanics* (Wiley, New York, 1966), Vol. I (see Chap. XI).
2 J. Dollard, J. Math. Phys. **5**, 729 (1964).
3 J. Dollard, Commun. Math. Phys. **12**, 193 (1969).
4 K. Kodaira, Amer. J. Math. **71**, 921 (1949).
5 D. Mulherin and I. I. Zinnes, J. Math. Phys. **11**, 1402 (1970).
6 Amrein, Martin, and Misra, Helv. Phys. Acta **43**, 313 (1970).
7 T. A. Green and O. F. Lanford, J. Math. Phys. **1**, 139 (1960).
8 S. T. Kuroda, J. Math. Phys. **3**, 933 (1962).
9 T. Kato, *Pertubation Theory of Linear Operators* (Springer, New York, 1966).
10 M. H. Stone, *Linear Transformations in Hilbert Space* (Amer. Math. Soc., New York, 1966), Vol. XV (see Chap. X).
11 M. O'Carroll, J. Math. Phys. 13, 268 (1972).
12 S. T. Kuroda, J. Math. Soc. Japan **11**, 247 (1959).
13 S. T. Kuroda, J. Math. Soc. Japan **12**, 243 (1960).
14 R. E. Edwards, *Fourier Series* (Holt, Reinhart, and Winston, New York, 1967), Vol. II.
15 E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second Order Differential Equations* (Oxford U.P., New York, 1962), Parts I and II.
16 N. N. Lebedev, *Special Functions and Their Applications* (Prentice-Hall, New York, 1965).
17 T. Ikebe, Arch. Ratl. Mech. Anal. 5, 1 (1960).
18 L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York)

On Weyl and Lyra Manifolds*

D. K. Sen and J. R. Vanstone

Department of Mathematics, University of Toronto, Toronto, Canada

(Received 1 November 1971; Revised Manuscript Received 6 February 1972)

It is shown that Weyl's geometry and an apparently similar geometry suggested by Lyra are special cases of manifolds with more general connections. The difference between the two geometries and their relationship with Riemannian geometry are clarified by giving a global formulation of Lyra's geometry. Finally the outline of a field theory based on the latter geometry is given.

1. INTRODUCTION

Shortly after Einstein's general theory of relativity Weyl[1] suggested the first so-called unified field theory based on a generalization of Riemannian geometry. In retrospect, it would seem more appropriate to call Weyl's theory a geometrized theory of gravitation and electromagnetism (just as the general theory was a geometrized theory of gravitation only), rather than a unified field theory. It is not quite clear to what extent the two fields have been unified, even though they acquire (different) geometrical significances in the same geometry. The theory was never taken seriously because it was based on the concept of nonintegrability of length transfer, and, as pointed out by Einstein, this implies that spectral frequencies of atoms depend on their past histories and therefore have no absolute significance. Nevertheless, Weyl's geometry provides an interesting example of non-Riemannian connections, and recently Folland[2] has given a global formulation of Weyl manifolds thereby clarifying considerably many of Weyl's basic ideas.

In 1951 Lyra[3] suggested a modification of Riemannian geometry which bears a remarkable resemblance to Weyl's geometry. But in Lyra's geometry, unlike Weyl's, the connection is metric preserving as in Riemannian geometry; in other words, length transfers are integrable. Lyra also introduced the notion of a gauge and in the "normal" gauge the curvature scalar is identical to that of Weyl. It is thus possible[4] to construct a geometrized theory of gravitation and electromagnetism much along the lines of Weyl's "unified" field theory without, however, the inconvenience of nonintegrable length transfer.

The original formulation by Lyra was couched in local coordinates (or rather what Lyra termed "reference systems") and thus the geometric significance of his definition of connection was somewhat unclear. In this paper we shall give an invariant formulation of Lyra manifolds and show how both the Weyl and Lyra manifolds can be regarded as special cases of manifolds with more general connections. We believe that we have thus clarified to a great extent the essential difference between the two geometries and their relationship with Riemannian geometry.

2. LINEARLY CONNECTED MANIFOLDS

M will always denote an n-dimensional C^∞ manifold, $T_m(M)$ the tangent space to M at $m \in M$, $\mathfrak{F}(M)$ the ring of C^∞ functions on M, $\mathfrak{X}(M)$ the Lie algebra of C^∞ vector fields on M, and $\Lambda^1(M)$ the $\mathfrak{F}(M)$-module of C^∞ 1-forms on M.

We shall suppose that M is endowed with a nonsingular metric, that is, a second-order symmetric covariant tensor field g such that, at every point $m \in M$, the induced form g_m on $T_m(M) \times T_m(M)$ is nondegenerate.

Recall that a linear connection ∇ on M is a mapping $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, written $(X, Y) \to \nabla_X Y$, such that (i) $\nabla_{f\cdot X + g\cdot Y}(Z) = f\cdot\nabla_X Z + g\cdot\nabla_Y Z$, (ii) $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$, (iii) $\nabla_X(f\cdot Y) = X(f)Y + f\nabla_X Y$ for all $f, g \in \mathfrak{F}(M), X, Y, Z \in \mathfrak{X}(M)$.

The torsion of ∇ is the mapping $\mathrm{Tor}_\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, given by $\mathrm{Tor}_\nabla(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$. A connection ∇ enables one to define a general covariant derivative of any tensor field with respect to a vector field Z, which preserves tensor types. For example, $\nabla_Z g$ is again a second-order covariant tensor field such that

$$\nabla_Z g(X, Y) = Z(g(X, Y)) - g(\nabla_Z X, Y) - g(\nabla_Z Y, X)$$

for all $X, Y \in \mathfrak{X}(M)$.

We begin with a general result regarding connections.[5]

Proposition 1: Let A and B be covariant tensor fields on M such that (i) $(Z, X, Y) \to A(Z, X, Y) \in \mathfrak{F}(M)$ is symmetric and (ii) $(X, Y) \to B(X, Y) \in \mathfrak{X}(M)$ is antisymmetric, in (X, Y). Then there exists a unique connection ∇ on M such that

$$\begin{aligned} \nabla_Z g(X, Y) &= A(Z, X, Y), \\ \mathrm{Tor}_\nabla(X, Y) &= B(X, Y) \end{aligned} \tag{2.1}$$

for all $X, Y, Z \in \mathfrak{X}(M)$.

Proof: (2.1) implies

$$\begin{aligned} g(\nabla_Z X, Y) = Z(g(X, Y)) - A(Z, X, Y) - g(B(Z, Y), X) \\ - g([Z, Y], X) - g(\nabla_Y Z, X). \end{aligned} \tag{2.2}$$

Let (2.2′) and (2.2″) denote the two more equations obtained by cyclic permutation of X, Y, and Z in (2.2). If we add (2.2′) and (2.2″), and subtract (2.2), we get

$$\begin{aligned} 2g(\nabla_X Y, Z) = X(g(Y, Z)) + Y(g(Z, X)) - Z(g(X, Y)) \\ - A(X, Y, Z) - A(Y, Z, X) + A(Z, X, Y) \\ - g(B(X, Z), Y) - g(B(Y, X), Z) \\ + g(B(Z, Y), X) - g([X, Z], Y) \\ - g([Y, X], Z) + g([Z, Y], X). \end{aligned} \tag{2.3}$$

Since g is nonsingular, this defines $\nabla_X Y$ uniquely. Conversely ∇, defined by (2.3), satisfies (2.1). QED

A connection ∇ is said to be *metric preserving* or to have *integrable length transfer* in the case $A = 0$—i.e. $\nabla_X g = 0$ for all $X \in \mathfrak{X}(M)$—and *torsion free* in the case $B = 0$—i.e., $\mathrm{Tor}_\nabla = 0$.

The special case of $A = B = 0$ is the Riemannian one; that is, a Riemannian connection is both torsion free and one for which length transfer is integrable.

In local coordinates (x^α), let $\{e_\alpha = \partial/\partial x^\alpha\}$ $(\alpha = 1, \ldots, n)$ be the associated coordinate vector fields. The C^∞ functions $\Gamma^\mu_{\alpha\beta}$ (components of affine connection), defined by $\nabla_{e_\beta} e_\alpha = \Gamma^\mu_{\alpha\beta} e_\mu$, are then, in the Riemannian case, simply the Christoffel symbols $\left\{ {\mu \atop \alpha\beta} \right\}$ constructed from the components of the metric tensor $g_{\alpha\beta} = g(e_\alpha, e_\beta)$. This is easily seen by putting $A = B = 0$, $X = e_\alpha$, $Y = e_\beta$, $Z = e_\gamma$ in (2.3) and noting that all Lie brackets vanish.

3. WEYL AND LYRA MANIFOLDS

A. Weyl Manifold

A Weyl manifold M (with metric g) is characterized by a 1-form $\phi \in \Lambda^1(M)$ and the Weyl connection is a special case of (2.1) with

$$\begin{aligned} A(Z, X, Y) &= -\phi(Z)\cdot g(X, Y), \\ B(X, Y) &= 0. \end{aligned} \tag{3.1}$$

The Weyl connection is thus torsion free but not metric preserving.

Proposition 2: In local coordinates the Weyl connection components are given by

$$\Gamma^\mu_{\alpha\beta} = \left\{ {\mu \atop \alpha\beta} \right\} + \tfrac{1}{2}(\delta^\mu_\alpha \phi_\beta + \delta^\mu_\beta \phi_\alpha - g_{\alpha\beta}\phi^\mu), \tag{3.2}$$

where $\phi_\alpha = \phi(e_\alpha)$ and $\phi^\mu = g^{\mu\alpha}\phi_\alpha$.

Proof: This is easily seen by substituting (3.1) in (2.3) and putting $X = e_\alpha$, $Y = e_\beta$, $Z = e_\gamma$. QED

The Weyl contracted curvature scalar computed from (3.2) is

$$K = R + \tfrac{1}{4}(n-2)(n-1)\,\phi_\alpha\phi^\alpha + (n-1)\,\phi^\alpha_{;\alpha}, \tag{3.3}$$

where R is the Riemannian curvature scalar and $\phi^\alpha_{;\alpha}$ the Riemannian covariant divergence of ϕ^α.

We recapitulate briefly Weyl's concept of gauge and length transfer as formulated by Folland.[2] We have seen that a Weyl manifold is specified by a metric g and a 1-form ϕ. Consider now the *equivalence class* $\{g\}$ of metrics strictly conformal to g, i.e., $g' \in \{g\}$ iff $g' = e^\lambda g$, $\lambda \in \mathfrak{F}(M)$. Each λ determines a *gauge* on the Weyl manifold and the 1-form ϕ determines a *Weyl structure*, namely, a mapping $F : \{g\} \to \Lambda^1(M)$, as follows: $F(e^\lambda g) = \phi - d\lambda$. Under a *gauge transformation* $g \to g' = e^\lambda g$, we have $\phi \to \phi' = \phi - d\lambda$.

A Weyl structure enables one to define the concept of length transfer as follows. Let $\gamma : [0, 1] \to M$ be a curve with $\gamma(0) = p, \gamma(1) = q$ and $g' \in \{g\}$. Then Folland has shown that

the (Weyl) parallel translate of $(g')_p$ along γ to q
$$= \exp\left(\int_0^1 \gamma^*(F(g'))\right) g'_q. \qquad (3.4)$$

The length transfer is independent of gauge in the following sense: If $g'', g' \in \{g\}$ agree at p, then their parallel translates along γ agree at q. Thus, although the length transfer depends on the path joining p to q, it is possible to *compare* lengths of (parallel) vector fields at p and q.

If M is simply connected, then the length transfer is independent of path, that is to say, *integrable* if and only if the *length curvature* $\Omega = -d\phi$ vanishes (e.g., if ϕ is exact).

B. Lyra Manifold

A Lyra manifold is also characterized by a 1-form ϕ, but the Lyra connection is a special case of (2.1) with

$$A(Z,X,Y) = 0, \qquad B(X,Y) = \tfrac{1}{2}[\phi(Y)X - \phi(X)Y]. \qquad (3.5)$$

The Lyra connection is thus metric preserving but not torsion free.[6] From (3.5) and (2.3) we have, for the Lyra connection,

$$2g(\nabla_X Y, Z) = X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) - g([X,Z],Y) - g([Y,X],Z) + g([Z,Y],X) - \phi(Z)g(X,Y) + \phi(Y)g(X,Z). \qquad (3.6)$$

The notion of gauge and "reference systems" introduced by Lyra can be made more precise as follows. By a *local reference system* on M we shall mean a triple (U_i, ψ_i, f_i), where U_i is a coordinate neighborhood of M, ψ_i the associated coordinate map [$\{(U_i, \psi_i)\}$ being an atlas of M], and $f_i: U_i \to \mathbb{R} - \{0\}$ a (nonzero) C^∞ function on U_i. The n coordinates (x_i^α) together with the gauge function $x_i^0 = f_i \circ \psi_i^{-1}$ characterize the local reference system (= coordinate system + gauge) in local coordinates. We shall therefore denote a local reference system also by $(x^0; x^\alpha)$. A local reference system induces a natural basis $\{\tilde{e}_\alpha(m) = [(x^0)^{-1}\partial/\partial x^\alpha]_m\}$ in $T_m(M)$ at $m \in U_i$ and a corresponding set of reference vector fields $\{\tilde{e}_\alpha = (x^0)^{-1}\partial/\partial x^\alpha\}$. Let (U'_j, ψ'_j, f'_j) be another local reference system such that $U_i \cap U'_j \neq \phi$. Then under a transformation of local reference systems $(x^0; x^\alpha) \to (x^{0'}; x^{\alpha'})$, the reference vector fields transform as follows:

$$\tilde{e}_{\alpha'} = \lambda^{-1} A^\alpha_{\alpha'} \tilde{e}_\alpha, \qquad (3.7)$$

where $\lambda = x^{0'}/x^0$ and $A^\alpha_{\alpha'} = \partial x^\alpha / \partial x^{\alpha'}$. Consequently, the components of a vector field X in local reference systems transform as follows ($X = X^\alpha \tilde{e}_\alpha$):

$$X^{\alpha'} = \lambda A^{\alpha'}_\alpha X^\alpha. \qquad (3.8)$$

It should be noted that the Lie Brackets of the reference vector fields do not vanish; instead, we have

$$[\tilde{e}_\alpha, \tilde{e}_\beta] = \tfrac{1}{2}(\delta^\mu_\alpha \mathring{\phi}_\beta - \delta^\mu_\beta \mathring{\phi}_\alpha)\tilde{e}_\mu, \qquad (3.9)$$

where $\mathring{\phi}_\alpha = -2\partial_\alpha(1/x^0)$.

The natural basis [dual to $\{\tilde{e}_\alpha(m)\}$] in $T^*_m(M)$ is $\{\tilde{e}^\alpha(m) = (x^0 dx^\alpha)_m\}$; and hence, if we denote the components of the metric tensor g, in a local reference system, by $g_{\alpha\beta} = g(\tilde{e}_\alpha, \tilde{e}_\beta)$, the classical metric form is given by $ds^2 = (x^0)^2 g_{\alpha\beta} dx^\alpha dx^\beta$. Denote the components of the Lyra connection ∇ in a local reference system $(x^0; x^\alpha)$ by $\tilde{\Gamma}^\mu_{\alpha\beta}$, i.e., $\nabla_{\tilde{e}_\beta} \tilde{e}_\alpha = \tilde{\Gamma}^\mu_{\alpha\beta} \tilde{e}_\mu$. We now show that ∇ is indeed the connection used by Lyra in a local reference system.

Proposition 3: In a local reference system the Lyra connection components are given by[7]

$$\tilde{\Gamma}^\mu_{\alpha\beta} = (1/x^0)\left\{{}^{\mu}_{\alpha\beta}\right\} + \tfrac{1}{2}(\delta^\mu_\beta \phi_\alpha - g_{\alpha\beta}\phi^\mu), \qquad (3.10)$$

where
$$\phi_\alpha = \phi(\tilde{e}_\alpha) + \mathring{\phi}_\alpha \qquad (3.11)$$
and
$$\phi^\mu = g^{\mu\alpha}\phi_\alpha.$$

Proof: Put $X = \tilde{e}_\alpha$, $Y = \tilde{e}_\beta$, $Z = \tilde{e}_\gamma$ in (3.6). Then, using (3.9), we get

$$2\tilde{\Gamma}^\mu_{\beta\alpha} g_{\mu\gamma} = (1/x^0)(\partial_\alpha g_{\beta\gamma} + \partial_\beta g_{\gamma\alpha} - \partial_\gamma g_{\alpha\beta}) + (\phi(\tilde{e}_\beta) + \mathring{\phi}_\beta) g_{\alpha\gamma} - (\phi(\tilde{e}_\gamma) + \mathring{\phi}_\gamma) g_{\alpha\beta},$$

which implies (3.10), in view of (3.11). QED

Note that, according to (3.11), the ϕ_α do not transform as covariant vector field components under a transformation of local reference systems, i.e., as in (3.7). From (3.11) it is easily seen that the transformation formula for ϕ_α is precisely the one given by Lyra:

$$\phi_\alpha \to \phi_{\alpha'} = \lambda^{-1} A^\alpha_{\alpha'} \{\phi_\alpha + (x^0)^{-1}\partial_\alpha(\ln\lambda^2)\}. \qquad (3.12)$$

Note also that, if ϕ is exact, there exists a local reference system in which $\phi_\alpha = 0$.

It is now simple to derive the transformation formula for $\tilde{\Gamma}^\mu_{\alpha\beta}$, as given by Lyra, from $\tilde{\Gamma}^{\mu'}_{\alpha'\beta'} \tilde{e}_{\mu'} = \nabla_{\tilde{e}_{\beta'}} \tilde{e}_{\alpha'} = \nabla_{\lambda^{-1} A^\beta_{\beta'} e_\beta} \lambda^{-1} A^\alpha_{\alpha'} e_\alpha$.

The Lyra contracted curvature scalar computed from (3.10) is

$$K = R/(x^0)^2 + \tfrac{1}{4}(n-2)(n-1)\phi^\alpha(\phi_\alpha + \mathring{\phi}_\alpha) + [(n-1)/x^0]\phi^\alpha_{;\alpha}, \qquad (3.13)$$

which reduces to the Weyl curvature scalar (3.3), if we choose the "normal" gauge $x^0 = 1$.

A geodesic of the metrical connection is an extremal curve $x^\mu = x^\mu(t)$ given by

$$\delta \int ds = \delta \int \left((x^0)^2 g_{\mu\lambda} \frac{dx^\mu}{dt}\frac{dx^\lambda}{dt}\right)^{1/2} dt = 0. \qquad (3.14)$$

The Euler–Lagrange equations for the geodesics turn out[8] to be (taking $t = s$)

$$\frac{d^2x^\mu}{ds^2} + \left\{{}^{\mu}_{\alpha\beta}\right\}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} + \frac{x^0}{2}(\delta^\mu_\alpha \mathring{\phi}_\beta + \delta^\mu_\beta \mathring{\phi}_\alpha - g_{\alpha\beta}\mathring{\phi}^\mu) \times \frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0. \qquad (3.15)$$

On the other hand, in view of (3.10), the autoparallel curves are given by

$$x^0 \frac{d^2x^\mu}{ds^2} + \left[\frac{1}{x^0}\left\{{\mu \atop \alpha\beta}\right\} + \tfrac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu)\right] \times (x^0)^2 \frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = \tfrac{1}{2}(\phi_\alpha - \mathring{\phi}_\alpha)(x^0)^2 \frac{dx^\alpha}{ds}\frac{dx^\mu}{ds}. \quad (3.16)$$

Thus, in Lyra's geometry, in contrast to the situation in the Riemannian case, the autoparallels associated with the affine connection do not coincide with the geodesics which arise from the metric.

4. A GEOMETRIZED FIELD THEORY

It is now possible to construct a unified field theory in the framework of Lyra's geometry with almost identical results as in Weyl's theory. We present here only a brief outline of the theory and refer to Refs. 4, 9, and 10 for details.

A variational principle for the field equations in Lyra's geometry would have the form

$$\delta \int \mathcal{W}\, dv = \delta \int W\sqrt{-g}\, x^0 dx^1 \cdots x^0 dx^4 = 0, \qquad (4.1)$$

where W is an absolute scalar function of $g_{\mu\lambda}$ and ϕ_μ, whose explicit form shall be considered later. If we express (4.1) in the form

$$0 = \delta \int \mathcal{W}\, dv = \int (\mathcal{W}^{\alpha\beta}\delta g_{\alpha\beta} + \mathcal{W}^\alpha \delta\phi_\alpha)dv, \qquad (4.2)$$

then the field equations are

$$\left.\begin{aligned} \mathcal{W}^{\alpha\beta} &= 0 \\ \mathcal{W}^{\alpha} &= 0 \end{aligned}\right\}. \qquad (4.3)$$

The invariance of the action integral under both gauge and coordinate transformations (i.e., transformations of reference systems) implies two sets of identities

$$\left.\begin{aligned} \mathcal{W}^\alpha_{\beta;\alpha} + \tfrac{1}{2}\phi_\beta\partial_\alpha\mathcal{W}^\alpha - \tfrac{1}{2}\mathcal{W}^\alpha f_{\alpha\beta} &= 0 \\ \partial_\alpha\mathcal{W}^\alpha + \mathcal{W}^\alpha_\alpha + \tfrac{1}{2}\mathcal{W}^\alpha\phi_\alpha &= 0 \end{aligned}\right\}, \qquad (4.4)$$

where $f_{\alpha\beta} = \partial_\beta\phi_\alpha - \partial_\alpha\phi_\beta$.

A simple choice for W is to consider

$$\delta \int (K - \alpha\Phi_{\mu\lambda}\Phi^{\mu\lambda})\sqrt{-g}\, dv = 0, \qquad (4.5)$$

where

$$\Phi_{\mu\lambda} = (1/x^0)f_{\mu\lambda} - \tfrac{1}{2}(\mathring{\phi}_\mu\phi_\lambda - \mathring{\phi}_\lambda\phi_\mu) \qquad (4.6)$$

is the basic second order antisymmetric tensor in the theory and α a constant. The field equations take the following form in the normal gauge $x^0 = 1$:

$$\left.\begin{aligned} R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R + \tfrac{3}{2}\phi_\alpha\phi_\beta - \tfrac{3}{4}g_{\alpha\beta}\phi_\mu\phi^\mu + 8\pi\alpha E_{\alpha\beta} &= 0 \\ [\partial(f^{\alpha\beta}\sqrt{-g})/\sqrt{-g}\ \partial x^\beta] + \tfrac{3}{4}\alpha\phi^\alpha &= 0 \end{aligned}\right\}, \qquad (4.7)$$

where $E^\beta_\alpha = f^{\beta\nu}f_{\alpha\nu} - \frac{1}{4}\delta^\beta_\alpha f_{\mu\lambda}f^{\mu\lambda}$. Apart from constant factors and a cosmological term, these equations are identical with Weyl's field equations.

Thus a geometrized field theory based on Lyra's geometry has all the essential features of Weyl's theory without, however, its primary defect.

---

* Supported in part by the National Research Council of Canada.

1 H. Weyl, S.-B. Preuss. Akad. Wiss., Berlin, 465 (1918).

2 G. Folland, J. Diff. Geom. **4**, 145 (1970).

3 G. Lyra, Math. Z. **54**, 52 (1951); E. Scheibe, Math. Z. **57**, 65 (1952).

4 D. K. Sen, thesis (University of Paris, 1958).

5 N. J. Hicks, *Notes on Differential Geometry*, Van Nostrand Mathematical Studies # 3 (Van Nostrand, Princeton, N.J., 1965), p. 104.

6 It is interesting to note that Schrödinger [Space–Time Structure (Cambridge U.P., Cambridge, 1954), p. 6] considered the case of a nonsymmetric but metric preserving connection $\Gamma^\mu_{\alpha\beta}$ (in local coordinates) and showed that the antisymmetric part $\Gamma^\mu_{\underset{\vee}{\alpha\beta}}$ determines $\Gamma^\mu_{\alpha\beta}$ uniquely as follows:

$$\Gamma^\mu_{\alpha\beta} = \left\{{\mu \atop \alpha\beta}\right\} + g^{\mu\lambda}g_{\alpha\nu}\Gamma^\nu_{\underset{\vee}{\beta\lambda}} + g^{\mu\lambda}g_{\beta\nu}\Gamma^\nu_{\underset{\vee}{\alpha\lambda}} + \Gamma^\mu_{\underset{\vee}{\alpha\beta}}.$$

7 Note that Lyra set $\tilde{\Gamma}^\mu_{\alpha\beta} = \Gamma^\mu_{\alpha\beta} - \frac{1}{2}\delta^\mu_\alpha\phi_\beta$ so that, from (3.10), $\Gamma^\mu_{\alpha\beta} = (1/x^0)\left\{{\mu \atop \alpha\beta}\right\} + \frac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu)$. Lyra's $\Gamma^\mu_{\alpha\beta}$ should not be confused with the components of ∇ in a local coordinate system, i.e. a local reference system with the "normal" gauge $x^0 = 1$.

8 D. K. Sen, Can. Math. Bull. **3**, 255 (1960).

9 D. K. Sen, *Fields and/or Particles* (Academic, London, 1968), p. 84.

10 K. A. Dunn, thesis (University of Toronto, 1971).

Perturbations of Gibbs States. I. Grand Canonical Formalism for Self-Interacting Fermion Systems

D. A. Uhlenbrock*

Department of Mathematics, University of Wisconsin Madison, Wisconsin 53706

(Received 17 January 1972)

The perturbation series for the statistical operators of quantum statistical mechanics developed earlier is applied to provide a perturbation theory for the grand canonical Gibbs states of a self-interacting fermion system. Explicit formulas for an interaction Hamiltonian which is a polynomial in creation–annihilation operators are provided.

INTRODUCTION

The equilibrium (quantum) statistical mechanics of infinite systems has received a great deal of attention recently. Nevertheless, due to the complexity of the mathematical setup involved, the majority of the results derived in this context are qualitative and general. Detailed investigations of specific (continuous) systems are still scarce.

It seems plausible that for general types of interactions the quantities of interest will be nearly impossible to calculate accurately. Accordingly the following approach to deriving *approximate* expressions would appear reasonable. In the first place the infinite system of interest can be approximated by a system contained in a finite (i.e., compact) volume if the thermodynamic limit makes sense. This reduction does not really run counter to the recent trend of concentrating interest on infinite systems; rather, it would aim to make full use of the existence of such an idealization and all the conclusions drawn from it, while at the same time providing approximate calculable information about it. In the C^*-algebra framework this amounts to examining the locally normal states locally.

$$x^0 \frac{d^2 x^\mu}{ds^2} + \left[\frac{1}{x^0}\left\{ {\mu \atop \alpha\beta} \right\} + \tfrac{1}{2}(\delta^\mu_\alpha \phi_\beta + \delta^\mu_\beta \phi_\alpha - g_{\alpha\beta}\phi^\mu)\right] \times (x^0)^2 \frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = \tfrac{1}{2}(\phi_\alpha - \mathring{\phi}_\alpha)(x^0)^2 \frac{dx^\alpha}{ds}\frac{dx^\mu}{ds}. \quad (3.16)$$

Thus, in Lyra's geometry, in contrast to the situation in the Riemannian case, the autoparallels associated with the affine connection do not coincide with the geodesics which arise from the metric.

4. A GEOMETRIZED FIELD THEORY

It is now possible to construct a unified field theory in the framework of Lyra's geometry with almost identical results as in Weyl's theory. We present here only a brief outline of the theory and refer to Refs. 4, 9, and 10 for details.

A variational principle for the field equations in Lyra's geometry would have the form

$$\delta \int \mathcal{W}\, dv = \delta \int W\sqrt{-g}\, x^0 dx^1 \cdots x^0 dx^4 = 0, \quad (4.1)$$

where W is an absolute scalar function of $g_{\mu\lambda}$ and ϕ_μ, whose explicit form shall be considered later. If we express (4.1) in the form

$$0 = \delta \int \mathcal{W}\, dv = \int (\mathcal{W}^{\alpha\beta}\delta g_{\alpha\beta} + \mathcal{W}^\alpha \delta\phi_\alpha)dv, \quad (4.2)$$

then the field equations are

$$\left.\begin{aligned} \mathcal{W}^{\alpha\beta} &= 0 \\ \mathcal{W}^\alpha &= 0 \end{aligned}\right\}. \quad (4.3)$$

The invariance of the action integral under both gauge and coordinate transformations (i.e., transformations of reference systems) implies two sets of identities

$$\left.\begin{aligned} \mathcal{W}^\alpha_{\beta;\alpha} + \tfrac{1}{2}\phi_\beta \partial_\alpha \mathcal{W}^\alpha - \tfrac{1}{2}\mathcal{W}^\alpha f_{\alpha\beta} &= 0 \\ \partial_\alpha \mathcal{W}^\alpha + \mathcal{W}^\alpha_\alpha + \tfrac{1}{2}\mathcal{W}^\alpha \phi_\alpha &= 0 \end{aligned}\right\}, \quad (4.4)$$

where $f_{\alpha\beta} = \partial_\beta \phi_\alpha - \partial_\alpha \phi_\beta$.

A simple choice for W is to consider

$$\delta \int (K - \alpha \Phi_{\mu\lambda}\Phi^{\mu\lambda})\sqrt{-g}\, dv = 0, \quad (4.5)$$

where

$$\Phi_{\mu\lambda} = (1/x^0) f_{\mu\lambda} - \tfrac{1}{2}(\mathring{\phi}_\mu \phi_\lambda - \mathring{\phi}_\lambda \phi_\mu) \quad (4.6)$$

is the basic second order antisymmetric tensor in the theory and α a constant. The field equations take the following form in the normal gauge $x^0 = 1$:

$$\left.\begin{aligned} R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R + \tfrac{3}{2}\phi_\alpha\phi_\beta - \tfrac{3}{4}g_{\alpha\beta}\phi_\mu\phi^\mu + 8\pi\alpha E_{\alpha\beta} &= 0 \\ [\partial(f^{\alpha\beta}\sqrt{-g})/\sqrt{-g}\ \partial x^\beta] + \tfrac{3}{4}\alpha\phi^\alpha &= 0 \end{aligned}\right\}, \quad (4.7)$$

where $E^\beta_\alpha = f^{\beta\nu} f_{\alpha\nu} - \frac{1}{4}\delta^\beta_\alpha f_{\mu\lambda} f^{\mu\lambda}$. Apart from constant factors and a cosmological term, these equations are identical with Weyl's field equations.

Thus a geometrized field theory based on Lyra's geometry has all the essential features of Weyl's theory without, however, its primary defect.

* Supported in part by the National Research Council of Canada.

1 H. Weyl, S.-B. Preuss. Akad. Wiss., Berlin, 465 (1918).
2 G. Folland, J. Diff. Geom. **4**, 145 (1970).
3 G. Lyra, Math. Z. **54**, 52 (1951); E. Scheibe, Math. Z. **57**, 65 (1952).
4 D. K. Sen, thesis (University of Paris, 1958).
5 N. J. Hicks, *Notes on Differential Geometry*, Van Nostrand Mathematical Studies #3 (Van Nostrand, Princeton, N.J., 1965), p. 104.
6 It is interesting to note that Schrödinger [Space–Time Structure (Cambridge U.P., Cambridge, 1954), p. 6] considered the case of a nonsymmetric but metric preserving connection $\Gamma^\mu_{\alpha\beta}$ (in local coordinates) and showed that the antisymmetric part $\Gamma^\mu_{\check{\alpha\beta}}$ determines $\Gamma^\mu_{\alpha\beta}$ uniquely as follows:
$$\Gamma^\mu_{\alpha\beta} = \left\{ {\mu \atop \alpha\beta} \right\} + g^{\mu\lambda}g_{\alpha\nu}\Gamma^\nu_{\check{\beta\lambda}} + g^{\mu\lambda}g_{\beta\nu}\Gamma^\nu_{\check{\alpha\lambda}} + \Gamma^\mu_{\check{\alpha\beta}}.$$
7 Note that Lyra set $\tilde{\Gamma}^\mu_{\alpha\beta} = \Gamma^\mu_{\alpha\beta} - \frac{1}{2}\delta^\mu_\alpha\phi_\beta$ so that, from (3.10), $\Gamma^\mu_{\alpha\beta} = (1/x^0)\left\{ {\mu \atop \alpha\beta} \right\} + \frac{1}{2}(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu)$. Lyra's $\Gamma^\mu_{\alpha\beta}$ should not be confused with the components of ∇ in a local coordinate system, i.e. a local reference system with the "normal" gauge $x^0 = 1$.
8 D. K. Sen, Can. Math. Bull. **3**, 255 (1960).
9 D. K. Sen, *Fields and/or Particles* (Academic, London, 1968), p. 84.
10 K. A. Dunn, thesis (University of Toronto, 1971).

Perturbations of Gibbs States. I. Grand Canonical Formalism for Self-Interacting Fermion Systems

D. A. Uhlenbrock*

Department of Mathematics, University of Wisconsin Madison, Wisconsin 53706

(Received 17 January 1972)

The perturbation series for the statistical operators of quantum statistical mechanics developed earlier is applied to provide a perturbation theory for the grand canonical Gibbs states of a self-interacting fermion system. Explicit formulas for an interaction Hamiltonian which is a polynomial in creation–annihilation operators are provided.

INTRODUCTION

The equilibrium (quantum) statistical mechanics of infinite systems has received a great deal of attention recently. Nevertheless, due to the complexity of the mathematical setup involved, the majority of the results derived in this context are qualitative and general. Detailed investigations of specific (continuous) systems are still scarce.

It seems plausible that for general types of interactions the quantities of interest will be nearly impossible to calculate accurately. Accordingly the following approach to deriving *approximate* expressions would appear reasonable. In the first place the infinite system of interest can be approximated by a system contained in a finite (i.e., compact) volume if the thermodynamic limit makes sense. This reduction does not really run counter to the recent trend of concentrating interest on infinite systems; rather, it would aim to make full use of the existence of such an idealization and all the conclusions drawn from it, while at the same time providing approximate calculable information about it. In the C^*-algebra framework this amounts to examining the locally normal states locally.

In the perturbation set up of interest to us, the dynamical behavior is described in terms of an unperturbed local Hamiltonian H and a local perturbation P. The corresponding local states are characterized by their *statistical operators* $S^{\bar{P}}(\beta) = \exp[-\beta(H + P)]$, $\beta > 0$. In a recent paper[1] we developed a perturbation theory for the semigroup $\beta \to S^{\bar{P}}(\beta)$ which expresses $S^{\bar{P}}(\beta)$ as a trace norm convergent perturbation series in terms of $S(\beta) = \exp(-\beta H)$ and powers of P for suitably restricted perturbations P.

In this paper we wish to make use of this perturbation series for the statistical operator and to develop a corresponding expansion for the correlation functions, i.e., the local state specified by the statistical operator.

Thus we are concerned with a system of self-interacting fermions in a finite volume described in terms of the grand canonical formalism. To avoid bulk and to simplify the structure of the expansion we assume the interaction is a polynomial in the creation and annihilation operators. The road to generalization from there is obvious. Besides the above series development for $S^{\bar{P}}(\beta)$, the second essential ingredient of the present perturbation theory is a knowledge of the unperturbed Gibbs state to which the perturbed Gibbs state is reduced.

OUTLINE

Section 1 sets the stage by specifying a test function space and providing a description of the Fock and anti-Fock representations of the canonical anticommutation relations in terms of the exterior algebra of the test function space. This way of looking at these two representations is natural in connection with the examination of their relation to the fermion C^*-Clifford algebra. In explicit terms, the left and right regular representations of the latter are unitarily equivalent to simple tensor product constructions of the former. The quasifree representations of the C^*-Clifford algebra can then be studied from both of these starting points.

In Sec. 2 the unperturbed statistical semigroup is specified in terms of the second quantization of the usual free single particle Hamiltonian. A review of properties of the associated Gibbs states is followed by an examination of the unitary equivalences or quasi-equivalences existing between such states and the Fock or Central states.

Section 3, finally, deals with the perturbation of these Gibbs states. In particular, the reduction of the perturbed Gibbs state to a series of terms each involving only the unperturbed Gibbs state is carried out. Then the individual series terms are analyzed and calculated as far as possible. The results are summarized in formula (36) of Theorem 3.1. It is clear from formula (36″) that further development depends on a specialization of the interaction.

1. PRELIMINARIES

A. Test Functions

Let $\mathcal{T}$ be a separable complex (H)-space with typical scalars $a, b, \cdots$ and elements $f, g, \cdots$. The inner product $f, g \to \langle f|g\rangle$ is $\mathbb{C}$-linear in f. If $(e_k)_{k\in K}$ is a complex orthonormal base (CONB) of $\mathcal{T}$, let $(e_k^*)_{k\in K}$ be the dual base of the dual (H)-space $\mathcal{T}^*$. For $f \in \mathcal{T}$, f^* denotes the element of $\mathcal{T}^*$ given by $\mathcal{T} \ni g \to \langle g|f\rangle$. With $\langle f^*|g^*\rangle = \overline{\langle f|g\rangle}$, the map $f \to f^*$ is the standard anti-unitary isomorphism of $\mathcal{T}$ with $\mathcal{T}^*$. We assume that the index set K is totally ordered by a relation $<$.

In some of our applications, the following concrete spaces are relevant: Given a point serving as coordinate origin any ν-dimensional real affine space looks like $\mathbb{R}^\nu$. With $\mathbb{R}_+ := \{\xi \in \mathbb{R} \mid \xi > 0\}$ and $b \in \mathbb{R}_+^\nu$, the set $B := \{x \in \mathbb{R}^\nu \mid 0 \le \xi_\rho < b_\rho,\ 1 \le \rho \le \nu\}$ is the *box* (cornered at the origin) *with diagonal* b. Its volume is $|B| := \Pi b_\rho > 0$.

Let $\mathbb{Z} := \{0, \pm 1, \pm 2, \cdots\}$ with $0<(-1)<1<(-2)<\cdots$ as a possible total ordering. If $b\mathbb{Z}^\nu$: $\{bm \mid bm = (b_\rho m_\rho)_{1\le\rho\le\nu},\ b_\rho \in b,\ m_\rho \in \mathbb{Z}\}$, the *periodic box with diagonal* b, B_p is defined as the quotient group $\mathbb{R}^\nu/b\mathbb{Z}^\nu$. If μ is Lebesgue measure on $\mathbb{R}^\nu$, $\mu_B(\cdot) := \mu(B \cap \cdot)$ is the induced measure on B while μ_{B_p} is Haar measure on B_p normalized to $|B| =: |B_p|$. Then $\mathcal{T}_B := L^2_{\mathbb{C}}(\mathbb{R}^\nu, \mu_B)$ is clearly isomorphic to $L^2_{\mathbb{C}}(B_p, \mu_{B_p})$ by the map which takes a function (class) of $\mathcal{T}_B$, restricts to B, extends periodically, and then factors by $b\mathbb{Z}^\nu$. $K_B := (2\pi/b)\mathbb{Z}^\nu$ can be totally ordered via the lexicographic ordering induced by $<$ on $\mathbb{Z}$. For $k \in K_B$, $e_k^B(x) := |B|^{-1/2}\exp(ik\cdot x)$ provides a CONB $(e_k^B)_{k\in K_B}$ of $\mathcal{T}_B$. In case b and B remain fixed, the super and subscripts B will usually be suppressed.

B. Fock and Anti-Fock Representation of the Canonical Anticommutation Relations (CAR)

We choose to view the Fock and anti-Fock constructions as analytic completions of algebraic constructions which are easily described in terms of exterior algebra.

Given $\mathcal{T}$, $(e_k)_{k\in K}$ as in Sec. 1A let $\Lambda\mathcal{T} = \bigoplus_{m\ge 0}\Lambda^m\mathcal{T}$ be the exterior algebra of $\mathcal{T}$. In this direct sum $\Lambda^0\mathcal{T} \cong \mathbb{C}\cdot 1_{\Lambda\mathcal{T}}$, $\Lambda^1\mathcal{T} \cong \mathcal{T}$, and $\Lambda^m\mathcal{T}$ is the mth exterior power of $\mathcal{T}$ for $m \ge 2$ which is spanned by decomposable vectors $f_1 \wedge f_2 \wedge \cdots \wedge f_m$, where $f_1, f_2, \ldots, f_m$ are in $\mathcal{T}$.

The assignment

$$\langle f_1 \wedge f_2 \wedge \cdots \wedge f_m | g_1 \wedge g_2 \wedge \cdots \wedge g_n\rangle := \begin{cases} 0, & m \ne n \\ \det(\langle f_i|g_j\rangle)_{\substack{1\le i\le m,\\ 1\le j\le n}}, & m = n \end{cases} \tag{1}$$

extends uniquely to an inner product on $\Lambda\mathcal{T}$ which gives it a (Hausdorff) pre-Hilbert space structure. Its (H)-space completion $\overline{\Lambda\mathcal{T}}$ is the (H)-space direct sum of the completions $\overline{\Lambda^m\mathcal{T}}$ of $\Lambda^m\mathcal{T}$ and is called the *Fock space over* $\mathcal{T}$.

If $n = (k_1, \ldots, k_r)$ with $k_1 < \cdots < k_r$ any *finite* subset of K define $e_n := \bigwedge_{k\in n} e_k = e_{k_1} \wedge \cdots \wedge e_{k_r}$. The corresponding collection $(e_n)_{n\subset K,\,\text{finite}}$ yields a CONB of $\overline{\Lambda\mathcal{T}}$.

For $f \in \mathcal{T}$, the linear operator on $\Lambda\mathcal{T}$: $f_1 \wedge \cdots \wedge f_m \to f \wedge f_1 \wedge \cdots \wedge f_m$ extends continuously to a linear operator $c(f) \in \mathcal{B}_\infty(\overline{\Lambda\mathcal{T}})$. These "creation operators" are linear in f and satisfy the CAR:

$$\begin{aligned} &c(f)c(g) + c(g)c(f) = 0, \\ &c(f)c(g)^+ + c(g)^+c(f) = \langle f|g\rangle\cdot 1 \end{aligned} \tag{2}$$

and the Fock condition: $c(f)^{+}1 = 0,\ \forall f \in \mathcal{T}$.

The *Fock algebra over* $\mathcal{T}$ is the C^*-subalgebra $\mathfrak{F}(\mathcal{T})$ of $\mathcal{B}_\infty(\overline{\Lambda}\mathcal{T})$ with 1 generated by the operators $c(f)$ for $f \in \mathcal{T}$. It is irreducible, simple, antiliminary, and uniformly hyperfinite of type 2^n. All of its *-representations are isometric.[2-4]

If u is an element of the unitary group $\mathfrak{U}(\mathcal{T})$ of $\mathcal{T}$, the map $f_1 \wedge \cdots \wedge f_m \to uf_1 \wedge \cdots \wedge uf_m$ extends uniquely to an automorphism Λu of $\Lambda\mathcal{T}$. Since it is isometric and has the isometric inverse Λu^{-1}, Λu extends by continuity to a unitary operator $\overline{\Lambda} u \in \mathfrak{U}(\overline{\Lambda}\mathcal{T})$. $u \to \overline{\Lambda} u$ is a strongly continuous representation of $\mathfrak{U}(\mathcal{T})$ in $\mathfrak{U}(\overline{\Lambda}\mathcal{T})$. Hence, if h is self-adjoint on $\mathcal{T}$, the corresponding unitary group $\mathbb{R} \ni t \to \exp(-ith)$ maps into $\mathbb{R} \ni t \to \overline{\Lambda}\exp(-ith) := \exp(-itd\overline{\Lambda}h)$. $d\overline{\Lambda}$ is the *second quantization map* for single particle observables.[2] Note that

$$c(uf) = (\overline{\Lambda} u)\cdot c(f)\cdot(\overline{\Lambda} u)^{-1}. \tag{3}$$

If $\mathcal{T}$ is interchanged with its dual $\mathcal{T}^*$ the same construction leads to the exterior algebra $\Lambda\mathcal{T}^*$, the (H)-space $\overline{\Lambda}\mathcal{T}^*$ and the Fock-algebra $\mathfrak{F}(\mathcal{T}^*)$ over $\mathcal{T}^*$. The anti-unitary map $\mathcal{T} \ni f \to f^* \in \mathcal{T}^*$ sends the basis $(e_k)_{k\in K}$ into the dual basis $(e_k^*)_{k\in K}$ and extends uniquely to a conjugate linear anti-isomorphism of $\Lambda\mathcal{T}$ and $\Lambda\mathcal{T}^*$ which sends $f_1 \wedge \cdots \wedge f_m$ to $(f_1 \wedge \cdots \wedge f_m)^* = f_m^* \wedge \cdots \wedge f_1^*$. This anti-isomorphism in turn extends to the anti-unitary map J from $\overline{\Lambda}\mathcal{T}$ onto $\overline{\Lambda}\mathcal{T}^*$. Thus $\overline{\Lambda}\mathcal{T}^*$ is canonically identified with the dual of $\overline{\Lambda}\mathcal{T}$, and with $e_n^* := (e_n)^* = Je_n$ the collection $(e_n^*)_{n\in K,\mathrm{finite}}$ is the dual basis of $(e_n)_{n\in K,\mathrm{finite}}$ for $\overline{\Lambda}\mathcal{T}$.

With the aid of * the creation operators over $\mathcal{T}^*$, $c(f^*)$ can be "pulled back" to $\mathcal{T}$ by means of the definition $d(f) := c(f^*)^{+}$. Then $\mathcal{T} \ni f \to d(f) \in \mathcal{B}_\infty(\overline{\Lambda}\mathcal{T}^*)$ is $\mathbb{C}$-linear and satisfies the CAR:

$$d(f)\cdot d(g) + d(g)\cdot d(f) = 0,$$
$$d(f)\cdot d(g)^{+} + d(g)^{+}\cdot d(f) = \langle f|g\rangle 1 \tag{4}$$

together with the anti-Fock condition $d(f)1 = 0$ for $f \in \mathcal{T}$. This condition leads to the interpretation of $d(f)$ as the "destruction operator" corresponding to $f \in \mathcal{T}$ just as $c(f)$ was the creation operator for $f \in \mathcal{T}$.

The anti-Fock algebra over $\mathcal{T}$, which is $\mathfrak{F}^*(\mathcal{T})$, is defined to be the Fock-algebra over $\mathcal{T}^*$: $\mathfrak{F}^*(\mathcal{T}) = \mathfrak{F}(\mathcal{T}^*)$. We refer to the maps $f \to c(f)$ resp. $f \to d(f)$ as the *Fock-* resp. *anti-Fock-representation* of the CAR over $\mathcal{T}$.

We note that $d(f) = J \circ c(f)^{+} \circ J^{-1}$ and $d(f)^{+} = J \circ c(f) \circ J^{-1}$.

C. Relation between the Fock and Clifford Algebra Constructions

In Ref. 2 Shale and Stinespring realized the Fock representation over $\mathcal{T}$ as the holomorphic spinor representation of the Clifford-algebra over $\mathcal{T}$. Looking at what one might call the antiholomorphic spinor representation, one finds that the anti-Fock representation is just as much buried in the Clifford algebra construction. In a sense nothing else is. We outline without proof (to be given elsewhere) how the Clifford algebra construction is *spatially* isomorphic to a tensor product construction involving the Fock and anti-Fock constructions. Apart from this we wish to establish our notation.

Given the complex (H)-space $\mathcal{T}$, let $\mathcal{T}_r$ be the underlying real space, i.e., the real vector space obtained from $\mathcal{T}$ by restricting the scalars to $\mathbf{R}$ together with the inner product $f, g \to \mathrm{Re}\langle f|g\rangle$. The complexification $\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r$ is a complex (H) space in a natural way as is $\mathcal{T} \oplus \mathcal{T}^*$. The map $\mathcal{T} \oplus \mathcal{T}^* \ni f \oplus g^* \to (1/\sqrt{2})(1 \otimes f - i \otimes if) + (1/\sqrt{2})(1 \otimes g - i \otimes ig) \in \mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r$ extends by linearity to a unitary isomorphism of the (H)-spaces involved. Under this isomorphism the anti-unitary map $z \otimes f \to \bar{z} \otimes f$ on $\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r$ goes over into $f \oplus g^* \to g \oplus f^* \equiv (f \oplus g^*)^*$ on $\mathcal{T} \oplus \mathcal{T}^*$.

The real symmetric bilinear form $f_1, f_2 \to \mathrm{Re}\langle f_1|f_2\rangle$ on $\mathcal{T}_r$ has a complexification which under this isomorphism corresponds to the complex symmetric bilinear form $f_1 \oplus g_1^*,\ f_2 \oplus g_2^* \to \langle f_1 \oplus g_1^* | (f_2 \oplus g_2^*)^*\rangle$ on $\mathcal{T} \oplus \mathcal{T}^*$. When considering $\mathcal{T}_r$, $\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r$, or $\mathcal{T} \oplus \mathcal{T}^*$ for the purposes of constructing their Clifford algebras we shall take them to be equipped with these symmetric bilinear forms (equivalently, the associated quadratic forms).

If C_0 denotes the Clifford functor, it is seen that $C_0(\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r)$ is canonically isomorphic as a complex algebra to the complexification $\mathbb{C} \otimes_{\mathbb{R}} C_0(\mathcal{T}_r)$ of $C_0(\mathcal{T}_r)$. By the above discussion they are clearly also naturally isomorphic to $C_0(\mathcal{T} \oplus \mathcal{T}^*)$.

The anti-unitary maps on $\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r$ and $\mathcal{T} \oplus \mathcal{T}^*$ mentioned above extend uniquely to involutions on $C_0(\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r)$ and $C_0(\mathcal{T} \oplus \mathcal{T}^*)$, respectively, which correspond to each other under this isomorphism. They will be denoted by $*$ and make $C_0(\mathbb{C} \otimes_{\mathbb{R}} \mathcal{T}_r)$ and $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ into $*$-algebras.

If E_0 is the unique central state on $C_0(\mathcal{T} \oplus \mathcal{T}^*)$, then $F, G \to \langle F|G\rangle := E_0(FG^*)$ is an inner product on $C_0(\mathcal{T} \oplus \mathcal{T}^*)$. Let $C^2(\mathcal{T} \oplus \mathcal{T}^*)$ be its (H)-space completion. Then the left and right regular representations of $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ extend continuously to faithful $*$-representations L and R of $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ in $\mathcal{B}(C^2(\mathcal{T} \oplus \mathcal{T}^*))$ such that for $F, G \in C_0(\mathcal{T} \oplus \mathcal{T}^*)$, $L(F)G = F\cdot G = R(G)F$.

By setting $\|F\|_\infty := \|L(F)\|$, the operator bound of $L(F)$, $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ is a normed $*$-algebra with a C^*-completion $C(\mathcal{T} \oplus \mathcal{T}^*)$—the C^*-Clifford algebra of $\mathcal{T}$.

Multiplication by -1 on $\mathcal{T} \oplus \mathcal{T}^*$ extends uniquely to the automorphism $C_0(-1)$ of $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ and by continuity to the unitary $C^2(-1)$ on $C^2(\mathcal{T} \oplus \mathcal{T}^*)$.

Proposition 1.1: There exists a unique linear bijection of $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ onto $\Lambda(\mathcal{T}) \otimes \Lambda(\mathcal{T}^*)$ such that for $f_i \oplus g_i^* \in \mathcal{T} \oplus \mathcal{T}^*$ with $1 \le i \le m$,

$\prod_{i=1}^{m}(f_i \oplus g_i^*)$ is sent to

$$\left(\prod_{i=1}^{m}\{[c(f_i) + c(g_i)^{+}] \otimes 1 + \overline{\Lambda}(-1) \times \otimes [d(f_i) + d(g_i)^{+}]\}\right)(1 \otimes 1). \tag{5}$$

This vector space isomorphism is isometric and hence extends uniquely to a unitary isomorphism of $C^2(\mathcal{T} \oplus \mathcal{T}^*)$ with the (H)-space tensor product $\overline{\Lambda}(\mathcal{T}) \overline{\otimes} \overline{\Lambda}(\mathcal{T}^*)$.

Under this (H)-space isomorphism various maps of interest correspond to each other as follows: For $f, g \in \mathcal{T}$:

$$\begin{aligned}
L(f) &\sim c(f)\otimes 1+\overline{\Lambda}(-1)\otimes d(f),\\
L(g^*) &\sim c(g)^+\otimes 1+\overline{\Lambda}(-1)\otimes d(g)^+,\\
R(f) &\sim c(f)\cdot\overline{\Lambda}(-1)\otimes\overline{\Lambda}(-1)+1\otimes\overline{\Lambda}(-1)\cdot d(f), \quad (6)\\
R(g^*) &\sim \overline{\Lambda}(-1)\cdot c(g)^+\otimes\overline{\Lambda}(-1)+1\otimes d(g)^+\cdot\overline{\Lambda}(-1),\\
C^2(-1) &\sim \overline{\Lambda}(-1)\otimes\overline{\Lambda}(-1),\\
* &\sim \pi_{12}\circ(*\otimes(*)^{-1}),
\end{aligned}$$

where π_{12} permutes the factors of $\overline{\Lambda}(\mathcal{T})\mathbin{\overline{\otimes}}\overline{\Lambda}(\mathcal{T}^*)$, and $*:\overline{\Lambda}(\mathcal{T})\to\overline{\Lambda}(\mathcal{T}^*)$ is the canonical anti-unitary map extending the antilinear anti-automorphism of $\Lambda(\mathcal{T})$ onto $\Lambda(\mathcal{T}^*)$ generated by $*:\mathcal{T}\to\mathcal{T}^*$.

We note that the anticommutation relations for L read as

$$L(f_1\oplus g_1^*)\cdot L(f_2\oplus g_2^*)+L(f_2\oplus g_2^*)\cdot L(f_1\oplus g_1^*) = 2\langle f_1\oplus g_1^*|(f_2\oplus g_2^*)^*\rangle\cdot 1. \quad (7)$$

D. Quasifree Representations

If T is a complex linear operator on $\mathcal{T}$ with $-1\le T\le 1$, then $T_\pm=[\frac{1}{2}(1\pm T)]^{1/2}$ are well-defined nonnegative operators such that $T_+^2+T_-^2=1$ and $T_+^2-T_-^2=T$.

Proposition 1.2: For $f\in\mathcal{T}\subset\mathcal{T}\oplus\mathcal{T}^*$ put

$$F_T(f)=L((1/\sqrt{2})(T_++T_-)f) + R((1/\sqrt{2})(T_+-T_-)f)\cdot C^2(-1) \quad (8)$$

and

$$G_T(f)=R((1/\sqrt{2})(T_++T_-)f) - L((1/\sqrt{2})(T_+-T_-)f)\cdot C^2(-1). \quad (8')$$

Then $f\to F_T(f)$ and $f\to G_T(f)$ generate two mutually commutative representations of the C^*-Clifford algebra of $\mathcal{T}$ on $C^2(\mathcal{T}\oplus\mathcal{T}^*)$.

Under the isomorphism of Proposition 1.1 these maps transform as

$$\begin{aligned}
F_T(f) &= F_1(T_+f)+F_{-1}(T_-f)\\
&\sim\sqrt{2}c(T_+f)\otimes 1+\sqrt{2}\overline{\Lambda}(-1)\otimes d(T_-f), \quad (9)\\
G_T(f) &= F_1(T_-f)\cdot C^2(-1)+C^2(-1)\cdot F_{-1}(T_+f)\\
&\sim\sqrt{2}c(T_-f)\cdot\overline{\Lambda}(-1)\otimes\overline{\Lambda}(-1)\\
&\quad+\sqrt{2}\,1\otimes\overline{\Lambda}(-1)\cdot d(T_+f). \quad (9')
\end{aligned}$$

Remark: The F_T above is essentially the F_T of Shale and Stinespring. The commutativity of F_T and G_T follows directly from that of R and L. The equivalences with the tensor product forms follow from the table of equivalences given in Proposition 1.1. In this form such representations are well known.[4,5] We are interested in detailing the unitary equivalence which shows how they arise in the C^*-Clifford algebra framework.

The formula $F_1(f)\sim\tilde{F}_1(f)\otimes 1=\sqrt{2}\,c(f)\otimes 1$ [resp. $F_{-1}(f)\sim\overline{\Lambda}(-1)\otimes\tilde{F}_{-1}(f)=\overline{\Lambda}(-1)\otimes\sqrt{2}\,d(f)$] shows how F_1 is a direct sum of Fock representations (resp. F_{-1} a direct sum of anti-Fock representations). Note that $\tilde{F}_1$ represents the C^*-Clifford algebra while $c=(1/\sqrt{2})\,\tilde{F}_1$ represents the CAR on $\overline{\Lambda}\mathcal{T}$. The basic anticommutators differ by a factor of 2.

Finally the formula $L\sim c\otimes 1+\overline{\Lambda}(-1)\otimes d$ shows how c and d are the essential ingredients in the construction of (the regular representation of) the C^*-Clifford algebra. In this sense c and d are buried in $C(\mathcal{T}\oplus\mathcal{T}^*)$ and essentially nothing else is.

2. UNPERTURBED SITUATION

A. Unperturbed Gibbs Semigroup

Let $\beta\in\mathbb{R}_+$ have the interpretation of an inverse temperature. From now on we assume given a self-adjoint linear operator h with domain $\mathfrak{D}_h\subset\mathcal{T}$ (the single particle Hamiltonian with the chemical potential subtracted) which generates a Gibbs semigroup on $\mathcal{T}$. By definition this means that the semigroup $\mathbb{R}_+\ni\beta\to\exp(-\beta h)$ has $\exp(-\beta h)\in\mathfrak{B}_1(\mathcal{T})$, the set of traceclass operators on $\mathcal{T}$, for all $\beta\in\mathbb{R}_+$.[1] Clearly there then exists a CONB $(e_k)_{k\in K}$ of $\mathcal{T}$ consisting of eigenvectors of $h: he_k=\epsilon_k e_k$, $k\in K$. Also $T_\beta e_k=t_{\beta,k}e_k$ for $k\in K$, $\beta\in\mathbb{R}_+$ with $T_\beta=\tanh(\frac{1}{2}\beta h)$ and $t_{\beta,k}=\tanh(\frac{1}{2}\beta\epsilon_k)$.

According to Proposition 3 of Ref. 2 $\exp(-\beta h)$ is in $\mathfrak{B}_1(\mathcal{T})$ if and only if $1-T_\beta$ is and also if and only if $\exp(-\beta d\overline{\Lambda}h)$ is in $\mathfrak{B}_1(\overline{\Lambda}\mathcal{T})$. If we write $H=d\overline{\Lambda}h$ for the second quantized Hamiltonian on Fock space $\overline{\Lambda}\mathcal{T}$ corresponding to h on $\mathcal{T}$, then $\beta\to\exp(-\beta h)$ is a Gibbs semigroup precisely if $\beta\to S(\beta)=\exp(-\beta H)$ is one. Furthermore, if $\tilde{F}_1$ is as in the remark following Proposition 1.2, the generating functional E_{T_β} of F_{T_β} is equal to the grand canonical expectation value: $L\in C(\mathcal{T}\oplus\mathcal{T}^*)$,

$$E_{T_\beta}(L)=\langle F_{T_\beta}(L)1|1\rangle=\frac{\mathrm{Tr}_{\overline{\Lambda}\mathcal{T}}[\tilde{F}_1(L)\cdot\exp(-\beta H)]}{\mathrm{Tr}_{\overline{\Lambda}\mathcal{T}}[1\cdot\exp(-\beta H)]}. \quad (10)$$

As F_{T_β} is cyclic with cyclic vector 1 it is the GNS representation of E_{T_β}. E_{T_β} is the *Gibbs state of temperature* β^{-1} *on* $C(\mathcal{T}\oplus\mathcal{T}^*)$ corresponding to the single particle Hamiltonian h on $\mathcal{T}$.

Whenever it is convenient we put

$$\tilde{E}_{T_\beta}(A)=\langle F_{T_\beta}(\tilde{F}_1^{-1}(A))1|1\rangle=\frac{\mathrm{Tr}_{\overline{\Lambda}\mathcal{T}}[A\exp(-\beta H)]}{\mathrm{Tr}_{\overline{\Lambda}\mathcal{T}}[1\cdot\exp(-\beta H)]} \quad (10')$$

for $A\in\mathfrak{F}(\mathcal{T})$, the Fock algebra of $\mathcal{T}$, so that

$$\tilde{E}_{T_\beta}(\tilde{F}_1(L))=E_{T_\beta}(L).$$

B. Some Properties of the Gibbs States

It was shown in Ref. 2 that E_{T_β} is concentrated on the real (commutative) subalgebra of $C(\mathcal{T}\oplus\mathcal{T}^*)$ which is generated by 1 and $s_k=1-e_k\cdot e_k^*=e_k^*e_k-1$ for all $k\in K$.

Let $\Delta=\Pi_{k\in K}\{1,-1\}$ be the product of $|K|$ many copies of the discrete two point space $\{1,-1\}$ with the product topology and the usual Borel structure. If Δ_0 is the subset of those $\gamma=(\gamma_k)_{k\in K}\in\Delta$ with only finitely many $\gamma_k=-1$, then there is a unique discrete probability measure μ_β concentrated on Δ_0 which assigns the weight $\Pi_{k\in K}\frac{1}{2}(1+\gamma_k t_{\beta,k})$ to $\gamma=(\gamma_k)_{k\in K}\in\Delta_0$. μ_β is a product measure which is quasi-invariant with respect to the action of Δ_0 on Δ. Let C_γ for $\gamma\in\Delta$ be the automorphism of $C(\mathcal{T}\oplus\mathcal{T}^*)$ which extends $e_k\to e_k$, $e_k^*\to e_k^*$ for $\gamma_k=1$, and $e_k\to e_k^*$, $e_k^*\to e_k$ for $\gamma_k=-1$. Then if E_1 is the Fock state and $L\in C(\mathcal{T}\oplus\mathcal{T}^*)$,

$$E_{T_\beta}(L)=\int_\Delta d\mu_\beta(\gamma)\cdot E_1(C_\gamma(L)) \quad (11)$$

gives a decomposition of E_{T_β} into pure states.

Also for $L \in C_0(\mathrm{span}(e_{k_i}, e^*_{k_i}) | 1 \le i \le N < \infty)$ the formula

$$E_{T_\beta}(L) = E_0\left(L \cdot \prod_{i=1}^{N}(1 + t_{\beta,k_i} s_{k_i})\right) = \langle L | \prod_{i=1}^{N}(1 + t_{\beta,k_i} s_{k_i})\rangle \qquad (12)$$

reduces the evaluation of E_{T_β} on $C_0(\mathcal{T} \oplus \mathcal{T}^*)$ to calculations involving E_0. We note that $\lim_{\beta\to 0} E_{T_\beta} = E_0$ and $\lim_{\beta\to\infty} E_{T_\beta} = E_1$. As KMS states the E_{T_β} share all the properties of this class of states.

In the terminology of Powers and Størmer[6] E_{T_β} is a *gauge-invariant generalized free state*. Their parameter $A = A_\beta$ corresponds to $\frac{1}{2}(1 - T_\beta)$ and is of traceclass by our assumptions on h. We include E_0 with $A_0 \Leftrightarrow \frac{1}{2}(1 - T_0) = \frac{1}{2}$ and E_1 with $A_\infty \Leftrightarrow \frac{1}{2}(1 - T_\infty) = 0$ as limiting cases but reserve β to be in $(0, \infty) = \mathbb{R}_+$.

Proposition 2.1: Let T_β be as in Sec. 2A, then

(i) E_0, E_{T_β}, and E_1 are factor states.

(ii) E_0 is type II_1, E_{T_β} and E_1 are type I_∞. Only E_1 is pure.

(iii) The cyclic representations F_0, F_{T_β}, F_1 canonically associated with E_0, E_{T_β}, E_1 have commutants which are factors of type II_1, I_∞, and I_1, respectively.

(iv) E_0 is not quasi-equivalent to either E_{T_β} or E_1; E_{T_β} is unitarily equivalent to every $E_{T_{\beta'}}$; E_{T_β} is quasi-equivalent but not unitarily equivalent to E_1.

Proof: (Note: The numbered theorems etc., quoted in this proof are from Ref. 6) (i) is well known for E_0 and E_1 and follows from Theorem 5.1 for E_{T_β}. By Lemma 5.3(i) E_{T_β} is of type I and since it is infinite (ii) follows.

For (iii): Since A_β has a trivial kernel it is *elementary* (Def. 5.5) if and only if it is of the form $A_\beta = \frac{1}{2}1 + H_\beta$ with H_β self-adjoint and of class HS (Hilbert-Schmidt). Since A_β is of traceclass and hence HS, this can happen only if $A_\beta - H_\beta = \frac{1}{2}1$ is HS. As $\mathcal{T}$ is infinite dimensional $\frac{1}{2}1$ is not HS and consequently A_β is *not* elementary. Note, however, that A_∞ is elementary. By Lemma 5.6 the commutant of $F_{T_\beta}(C(\mathcal{T} \oplus \mathcal{T}^*))$ is infinite and, since it is type I, (iii) follows.

For (iv): Since E_0 is type II_1 and E_{T_β}, E_1 are type I_∞ by (ii), E_0 is not quasi-equivalent to either E_{T_β} or E_1. Consider $1 - (1 - A_\beta)^{1/2} = 1 - [\frac{1}{2}(1 + T_\beta)]^{1/2} \geq 0$.

The function $[-1, 1] \ni t \to 1 - [\frac{1}{2}(1 + t)]^{1/2}$ is nonnegative, decreasing, and bounded from above by $\frac{1}{2}(1 - t)$. Hence $0 \le 1 - (1 - A_\beta)^{1/2} \le \frac{1}{2}(1 - T_\beta) = A_\beta$ implies $1 - (1 - A_\beta)^{1/2}$ is HS, since A_β is even traceclass. Consequently for any $0 < \beta < \beta' < \infty$, A_β and $A_{\beta'}$ are not elementary, $A_\beta^{1/2} - A_{\beta'}{}^{1/2}$ is *HS*, and $(1 - A_\beta)^{1/2} - (1 - A_{\beta'})^{1/2} = [1 - (1 - A_{\beta'})^{1/2}] - [1 - (1 - A_\beta)^{1/2}]$ is HS. Then Theorem 5.7(1) shows that E_{T_β} is unitarily equivalent to $E_{T_{\beta'}}$.

Finally, since A_β is not elementary while A_∞ is elementary, Theorem 5.7(1) also yields that E_{T_β} is not unitarily equivalent to E_1. On the other hand, $A_\beta^{1/2} - A_\infty^{1/2} = A_\beta^{1/2}$ is HS and $(1 - A_\beta)^{1/2} - (1 - A_\infty)^{1/2} = (1 - A_\beta)^{1/2} - 1$ is HS so that Theorem 5.1 implies the quasi-equivalence of E_{T_β} and E_1. ///

3. PERTURBATION OF GIBBS STATES

A. Perturbation of Gibbs Semigroups

We briefly collect some needed results on the perturbation of statistical semigroups which we developed elsewhere.[1]

With $\overline{H} = -H$ of Sec. 2A we consider $\mathbb{R}_+ \ni \beta \to \exp(\beta\overline{H}) \in \mathcal{B}_1(\overline{\Lambda\mathcal{T}})$ as the unperturbed semigroup.

The following definition is adopted from Hille and Phillips and works for more general semigroups[7]:

Definition: A linear operator $\overline{P}$ with domain $\mathfrak{D}_{\overline{P}} \subset \overline{\Lambda\mathcal{T}}$ is in the *Phillips perturbing class* $\mathcal{P}(\overline{H})$ of $\overline{H}$, where $\overline{H}$ is the infinitesimal generator of $S(\cdot)$, if

(i) $\mathfrak{D}_{\overline{P}} = \mathfrak{D}_{\overline{H}}$.

(ii) $\overline{P}$ is $\overline{H}$ bounded, i.e., for suitable constants c_1, c_2: $\forall f \in \mathfrak{D}_{\overline{H}}$ or $[\|\overline{P}f\| \le c_1 \|\overline{H}f\| + c_2\|f\|$.

(iii) $\forall s > 0$: $n(s) = \sup_{\|f\|=1, f\in\mathfrak{D}_{\overline{H}}} \|P \circ S(s)f\| < \infty$.

(iv) $\int_0^1 ds\, n(s) < \infty$.

Remark: For symmetric $\overline{P} \subset \overline{P}^*$ with $\mathfrak{D}_{\overline{P}} = \mathfrak{D}_{\overline{H}}$, conditions (ii) and (iii) of this definition are automatically satisfied and all that is needed for condition (iv) is that $n(s)$ be integrable at 0_+.

Theorem 3.1: Let H with domain $\mathfrak{D}_{\overline{H}}$ generate the Gibbs semigroup $S(\cdot)$ on $\overline{\Lambda\mathcal{T}}$.
(a) Assume $\overline{P}$ is a symmetric perturbing operator in $\mathcal{P}(\overline{H})$. Then $\overline{H} + \overline{P}$ defined on $\mathfrak{D}_{\overline{H}}$ generates a norm-continuous positive (C_0)-semigroup $\mathbb{R}_+ \ni \beta \to S^{\overline{P}}(\beta) = \exp[\beta(\overline{H} + \overline{P})]$ on $\overline{\Lambda\mathcal{T}}$.

Also, $S^{\overline{P}}(\beta)$ is expressible as a sum

$$S^{\overline{P}}(\beta) = \sum_{n=0}^{\infty} S_n^{\overline{P}}(\beta), \qquad \text{with } S_0^{\overline{P}}(\beta) = S(\beta) \qquad (13)$$

and for

$$n \ge 1,\ \Delta_n^\beta = \{(\beta_1, \ldots, \beta_n) \in R^n | \times \beta \ge \beta_1 \ge \beta_2 \ge \cdots \ge \beta_n \ge 0\},$$
$$S_n^{\overline{P}}(\beta) = \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n S(\beta - \beta_1) \times \overline{P}S(\beta_1 - \beta_2)\overline{P} \cdots \overline{P}S(\beta_n). \qquad (14)$$

The integrals of Eq. (14) are convergent Bôchner integrals (in tracenorm), so that $S_n^{\overline{P}}(\beta)$ is of traceclass and tracenorm-continuous in β for $\beta > 0$.

Also, for arbitrary $A \in \mathcal{B}(\overline{\Lambda\mathcal{T}})$ the formula

$$\mathrm{Tr}_{\overline{\Lambda\mathcal{T}}}[AS_n^{\overline{P}}(\beta)] = \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \times \mathrm{Tr}_{\overline{\Lambda\mathcal{T}}}[AS(\beta - \beta_1)\overline{P}\cdot \ldots \cdot \overline{P}S(\beta_n)] \qquad (15)$$

holds with absolutely convergent integrals. The sum of Eq. (13) is absolutely convergent in operator norm and uniformly so on intervals of the form $(0, \beta_0)$ for $0 < \beta_0 < \infty$.

(b) If instead of (a) it is assumed that $\overline{P}$ is symmetric and bounded with $\mathfrak{D}_{\overline{P}} = \mathfrak{D}_{\overline{H}}$, then $P \in \mathcal{P}(\overline{H})$ and all the conclusions of part (a) hold together with:

The sum of Eq. (13) is even absolutely convergent in *tracenorm* so that $\beta \to S^{\bar{P}}(\beta)$ is again a Gibbs semigroup on $\bar{\Lambda}\mathcal{T}$.

Furthermore, beyond Eq. (15) the following formula for traces is valid with absolute convergence: $A \in \mathcal{B}(\bar{\Lambda}\mathcal{T})$:

$$\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[AS^{\bar{P}}(\beta)] = \sum_{n=0}^{\infty} \mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[AS_n^{\bar{P}}(\beta)]. \tag{16}$$

It is to be expected that the boundedness condition can be replaced by weaker assumptions if not dropped altogether. Also more can be said about the $\bar{P}$ dependence of the perturbation development. In the present circumstances this dependence is (real) analytic in λ if $\bar{P}$ is replaced by $\lambda\bar{P}$ with $\lambda \in \mathbb{R}$.

B. Reduction to the Unperturbed Gibbs State

We assume from now on that the situation described in Theorem 3.1(b) with a $\bar{P} \in \mathcal{B}(\bar{\Lambda}\mathcal{T})$ prevails.

The perturbed Gibbs state $E_{T_\beta}^{\bar{P}}$

$$C(\mathcal{T} \oplus \mathcal{T}^*) \ni L \to E_{T_\beta}^{\bar{P}}(L) = \frac{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[\tilde{F}_1(L)S^{\bar{P}}(\beta)]}{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[1 \cdot S^{\bar{P}}(\beta)]} \tag{17}$$

can then be calculated in terms of the series expansion for $\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[AS^{\bar{P}}(\beta)]$ provided by Eq. (16).

In terms of the ordered CONB $(e_k)_{k\in K}$ for $\mathcal{T}$ mentioned in Sec. 2A we let $\mathcal{T}_N$ be the (H)-subspace of $\mathcal{T}$ spanned by $(e_k)_{k\in K_N}$, where K_N is the segment of K containing the first N elements of K (N finite). Then $\cup_N \mathcal{T}_N$ is dense in $\mathcal{T}$ and the subalgebras $C(\mathcal{T}_N \oplus \mathcal{T}_N^*) = C_0(\mathcal{T}_N \oplus \mathcal{T}_N^*)$ are increasing with N and have a union which is dense in $C(\mathcal{T} \oplus \mathcal{T}^*)$.

If κ and κ' denote arbitrary finite subsets of K, we define

$$c_\kappa = \prod_{k\in\kappa} c(e_k) \quad \text{and} \quad c_{\kappa'}^* = (c_{\kappa'})^*, \tag{18}$$

where the order of the terms in the product is compatible with that of K. An expression like

$$\bar{P} = \sum_{\kappa\kappa'}{}' p_{\kappa\kappa'}\, c_\kappa c_{\kappa'}^* \tag{19}$$

with $p_{\kappa\kappa'} = \bar{p}_{\kappa'\kappa}$ and Σ' indicating a finite sum is then a self-adjoint polynomial in the creation and annihilation operators which is in $\tilde{F}_1(C_0(\mathcal{T}_N \oplus \mathcal{T}_N^*))$ for all large enough N. Any self-adjoint $\bar{P}$ in the Fock algebra can be approximated in norm by such polynomials. Since the perturbation development is continuous in $\bar{P}$ with respect to a certain (complete) metric on $\mathcal{P}(\bar{H})$ which is equivalent to the operator norm metric on $\mathcal{P}(\bar{H}) \cap \mathcal{B}(\bar{\Lambda}\mathcal{T}) = \mathcal{B}(\bar{\Lambda}\mathcal{T})$, it follows that results shown to hold for such polynomial perturbations will extend by continuity to more general perturbations in $\mathcal{P}(\bar{H})$.

In order to keep the complexity of the ensuing expansions down we therefore make the (easily improvable) assumption that $\bar{P}$ is of the form given in Eq. (19). This will allow us to get a feel for the terrain and appears even more reasonable in this statistical mechanics context than similar cut-off assumptions in relativistic quantum mechanics.

Lemma 3.1: The assignment $c(e_k) \to \exp(\beta\epsilon_k)c(e_k)$; $c(e_k^*) \to \exp(-\beta\epsilon_k)\cdot c(e_k^*)$ for $k \in K$ and $he_k = \epsilon_k e_k$ extends uniquely to an automorphism a_β of the polynomial algebra in the $c(e_k)$ and $c(e_k^*)$.

If $\bar{P}$ is as in Eq. (19), then

$$a_\beta(\bar{P}) \equiv \bar{P}(\beta) = \sum{}' p_{\kappa\kappa'} \exp[\beta(\epsilon_\kappa - \epsilon_{\kappa'})]\cdot c_\kappa c_{\kappa'}^*, \tag{20}$$

with $\epsilon_\kappa = \sum_{k\in\kappa} \epsilon_k$.

Furthermore, it follows that

$$S_n^{\bar{P}}(\beta) = S(\beta)\cdot \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \bar{P}(\beta_1) \cdots \bar{P}(\beta_n) \tag{21}$$

and therefore for $A = \tilde{F}_1(L)$, $L \in C(\mathcal{T} \oplus \mathcal{T}^*)$,

$$\frac{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[S^{\bar{P}}(\beta)A]}{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[S(\beta)]} = \sum_{n=0}^{\infty} \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \tilde{E}_{T_\beta}(\bar{P}(\beta_1) \times \cdots \bar{P}(\beta_n)A). \tag{22}$$

Remark: a_β is closely related to the automorphism $j_{2\beta}$ defined on the algebra $\tilde{\mathfrak{a}} \subset \mathfrak{A}$ in the notation of Kastler, Pool, and Poulsen.[8] Hence a_β clearly extends much beyond the polynomial algebra.

Proof: The existence and uniqueness of a_β is clear. Then $a_\beta(c_\kappa) = \exp(\beta\epsilon_\kappa)c_\kappa$ and $a_\beta(c_\kappa^*) = \exp(-\beta\epsilon_\kappa)c_\kappa^*$ follows.

It is not hard to show that $c(e_k)S(\beta) = \exp(\beta\epsilon_k)\cdot S(\beta)c(e_k) = S(\beta)a_\beta(c(e_k))$ and $c(e_k^*)S(\beta) = \exp(-\beta\epsilon_k)\cdot S(\beta)c(e_k) = S(\beta)a_\beta(c(e_k^*))$. Consequently $\bar{P}\,S(\beta) = S(\beta)\cdot a_\beta(\bar{P}) = S(\beta)\cdot\bar{P}(\beta)$, where $\bar{P}(\beta)$ is as in Eq. (20).

This identity together with Eq. (14) implies

$$S_n^{\bar{P}}(\beta) = \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n S(\beta - \beta_1)\bar{P} \times \cdots \bar{P}\,S(\beta_{n-1} - \beta_n)S(\beta_n)\bar{P}(\beta_n)$$

and by successively sweeping the S factors to the left

$$= \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n S(\beta - \beta_1)S(\beta_1)\bar{P}(\beta_1)\cdot \ldots \cdot \bar{P}(\beta_n),$$

which gives Eq. (21) by the semigroup property of S. By utilizing Eqs. (10), (15), (16), and (21) it is seen that

$$\mathrm{Tr}_{\bar{\Lambda}}\,[S^{\bar{P}}(\beta)A] = \sum_{n=0}^{\infty} \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \times \mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}(S(\beta)\bar{P}(\beta_1) \cdots \bar{P}(\beta_n)A)$$

which yields Eq. (22). ///

We conclude from Eq. (22) that the perturbed state $E_{T_\beta}^{\bar{P}}$ has the following reduction to the unperturbed state $E_{T_\beta}: A = \tilde{F}_1(L)$:

$$E_{T_\beta}^{\bar{P}}(L) = \frac{\sum_{n=0}^{\infty} \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \tilde{E}_{T_\beta}(\bar{P}(\beta_1)\cdot \cdots \cdot\bar{P}(\beta_n)\cdot A)}{\sum_{n=0}^{\infty} \int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \tilde{E}_{T_\beta}(\bar{P}(\beta_1)\cdot \cdots \cdot\bar{P}(\beta_n))}. \tag{23}$$

Clearly, if $\bar{P} \to \lambda\bar{P}$, $\lambda \in \mathbb{C}$ the expression is at least meromorphic in λ.

C. Evaluation of Series Terms

According to formula (23) the viability of the perturbation expansion depends primarily on the ability to calculate the series terms

$$\int_{\Delta_n^\beta} d\beta_1 \cdots d\beta_n \tilde{E}_{T_\beta}(\bar{P}(\beta_1)\cdot \ldots \cdot \bar{P}(\beta_n)\cdot \tilde{F}_1(L)) \tag{24}$$

in reasonably economical ways.

We assume a fixed N large enough, so that $\overline{P} \in \mathfrak{F}_N \equiv \tilde{F}_1(C(\mathcal{T}_N \oplus \mathcal{T}_N^*)) \subset \mathfrak{F}(\mathcal{T})$. By the continuity and factor state properties of $\tilde{E}_{T_\beta}$ the terms corresponding to $\tilde{F}_1(L) \in \mathfrak{F}_N$ are of greatest import. Hence it will be convenient to use the well-known $*$-isomorphism of $\mathfrak{F}_N$ with a full matrix algebra of dimension $(2^N)^2$. To that end we define for $k \in K_N$

$$\begin{aligned} m_{00}^k &= c(e_k)^* c(e_k), \\ m_{11}^k &= c(e_k)\cdot c(e_k)^*, \\ m_{10}^k &= c(e_k)\cdot \prod_{l<k}[c(e_l)^*c(e_l) - c(e_l)c(e_l)^*], \\ m_{01}^k &= c(e_k)^* \cdot \prod_{l<k}[c(e_l)^*c(e_l) - c(e_l)c(e_l)^*]. \end{aligned} \tag{25}$$

If $\mathbb{Z}_2 = \{0,1\}$ is the cyclic group of order two, we denote by $\sigma = (\sigma_k)_{k\in K_N}$ an element of $(\mathbb{Z}_2)^N$, where $\sigma_k \in \mathbb{Z}_2$ for $k \in K_N$. Given $\sigma, \tau \in (\mathbb{Z}_2)^N$, put

$$M_{\sigma\tau} = \prod_{k\in K_N} m_{\sigma_k \tau_k}^k. \tag{26}$$

Lemma 3.2: The family $(M_{\sigma\tau})$, $\sigma, \tau \in (\mathbb{Z}_2)^N$ constitutes a full system of matrix units for the algebra $\mathfrak{F}_N$, i.e., $\mathfrak{F}_N$ is spanned by the $M_{\sigma\tau}$ and

$$M_{\sigma\tau} M_{\sigma'\tau'} = \delta_{\tau\sigma'} M_{\sigma\tau'}. \tag{27}$$

Furthermore, the defining equations (25) are inverted by

$$\begin{aligned} c(e_k) &= m_{10}^k \cdot \prod_{l\neq k}^{N} [m_{00}^l + \operatorname{sgn}(l-k)\cdot m_{11}^l], \\ c(e_k)^* &= m_{01}^k \cdot \prod_{l\neq k}^{N} [m_{00}^l + \operatorname{sgn}(l-k)\cdot m_{11}^l]. \end{aligned} \tag{28}$$

Equations (26) and (29) imply that $c(e_k)$ and $c(e_k)^*$ are linear combinations of the $M_{\sigma\tau}$ with $\pm$ 1-coefficients. Let $\overline{P} \in \mathfrak{F}_N$ have the expansion

$$\overline{P} = \sum_{\sigma,\tau\in\mathbb{Z}_2^N} \overline{P}_{\sigma\tau}\cdot M_{\sigma\tau}, \qquad \overline{P}_{\sigma\tau} = \overline{P}_{\tau\sigma}^* \in \mathbb{C}. \tag{29}$$

According to Lemma 3.1 this leads to

$$\begin{aligned} \overline{P}(\beta) &= a_\beta\Big(\sum_{\sigma,\tau} \overline{P}_{\sigma\tau} M_{\sigma\tau}\Big) \\ &= \sum_{\sigma,\tau} \overline{P}_{\sigma\tau}\, a_\beta(M_{\sigma\tau}) \\ &= \sum_{\sigma,\tau} \overline{P}_{\sigma\tau}(\beta)\cdot M_{\sigma\tau}, \end{aligned} \tag{30}$$

with the abbreviations

$$\sigma\cdot\epsilon = \sum_{k\in K_N} \sigma_k\cdot\epsilon_k, \qquad \tau\cdot\epsilon = \sum_{k\in K_N} \tau_k\cdot\epsilon_k, \text{ and}$$

$$\overline{P}_{\sigma\tau}(\beta) = \overline{P}_{\sigma\tau}\cdot\exp[\beta(\sigma\cdot\epsilon - \tau\cdot\epsilon)]. \tag{31}$$

Lemma 3.3: Let $\beta > 0$, $\sigma, \tau \in \mathbb{Z}_2^N$; then

$$\tilde{E}_{T_\beta}(M_{\sigma\tau}) = \delta_{\sigma\tau}\cdot \bar{n}_\sigma(\beta) \tag{32}$$

with

$$\bar{n}_\sigma(\beta) = \prod_{k\in K_N} \frac{\exp(-\beta\sigma_k\cdot\epsilon_k)}{1+\exp(-\beta\epsilon_k)}. \tag{32'}$$

Proof: By Eq. (26)

$$\tilde{E}_{T_\beta}(M_{\sigma\tau}) = \tilde{E}_{T_\beta}\Big(\prod_{k\in K_N} m_{\sigma_k\tau_k}^k\Big).$$

Only the $m_{\sigma_k\tau_k}^k$ with $\tau_k = \sigma_k$ belong to the real subalgebra of $\mathfrak{F}(\mathcal{T})$ generated by 1 and $\tilde{F}_1(s_k)$, $k \in K$. Since $\tilde{E}_{T_\beta}$ vanishes off this subalgebra (Sec. B), it follows that $\tilde{E}_{T_\beta}(M_{\sigma\tau}) = \delta_{\sigma\tau}\tilde{E}_{T_\beta}(M_{\sigma\sigma})$. By the factor state properties $\tilde{E}_{T_\beta}(M_{\sigma\sigma}) = \tilde{E}_{T_\beta}(\prod_{k\in K_N} m_{\sigma_k\sigma_k}^k) = \prod_{k\in K_N} \tilde{E}_{T_\beta}(m_{\sigma_k\sigma_k}^k)$. Since, as is easily verified with the aid of Eq. (12), $\tilde{E}_{T_\beta}(m_{00}^k) = [1+\exp(-\beta\epsilon_k)]^{-1}$ and $\tilde{E}_{T_\beta}(m_{11}^k) = \exp(-\beta\epsilon_k)\cdot[1+\exp(-\beta\epsilon_k)]^{-1}$ the lemma is proved. ///

Equations (28), (31), and (32) lead to

$$\begin{aligned} &\int_{\Delta_n^\beta} d\beta_1\cdot \cdots \cdot d\beta_n\, \tilde{E}_{T_\beta}(M_{\sigma_0\tau_0}\overline{P}(\beta_1)\cdot \cdots \cdot \overline{P}(\beta_n)) \\ &= \int_{\Delta_n^\beta} d\beta_1\cdot \cdots \cdot d\beta_n \sum_{\substack{\sigma_1\cdots\sigma_n \\ \tau_1\cdots\tau_n}} \overline{P}_{\sigma_1\tau_1}(\beta_1)\cdot \cdots \cdot \overline{P}_{\sigma_n\tau_n}(\beta_n) \\ &\quad \times \tilde{E}_{T_\beta}(M_{\sigma_0\tau_0}\cdot \cdots \cdot M_{\sigma_n\tau_n}) \\ &= \int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n \sum_{\tau_1\cdots\tau_n} \delta_{\sigma_0\tau_n}\bar{n}_{\sigma_0}(\beta)\cdot\overline{P}_{\tau_0\tau_1}(\beta_1)\cdot \cdots \\ &\quad \cdot \overline{P}_{\tau_{n-1}\tau_n}(\beta_n) = \sum_{\tau_1\cdots\tau_n} \delta_{\sigma_0\tau_n}\bar{n}_{\sigma_0}(\beta)\cdot\overline{P}_{\tau_0\tau_1}\cdot \cdots \\ &\quad \cdot \overline{P}_{\tau_{n-1}\tau_n}\cdot\int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n \times \exp\Big(\sum_{j=1}^n \beta_j(\tau_{j-1}-\tau_j)\cdot\epsilon\Big). \end{aligned} \tag{33}$$

Lemma 3.4: If $(\alpha_i)_{1\le i\le n}$ are complex numbers and $\gamma_0 = 0, \gamma_j = \Sigma_{i=1}^j \alpha_i$, $1 \le j \le n$, then the integral

$$\mathcal{T}_n(\beta;\alpha_1,\ldots,\alpha_n) = \int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n \exp\Big(\sum_{i=1}^n \alpha_i\beta_i\Big)$$

is equal to the sum of the residues of the meromorphic function

$$\mathbb{C} \ni s \to \exp(s\beta)\cdot\prod_{j=0}^n (s-\gamma_j)^{-1}.$$

In particular, if the $(\gamma_j)_{0\le j\le n}$ are all distinct, so that this function has only first order poles, then

$$\mathcal{T}_n(\beta;\alpha_1\cdots\alpha_n) = \sum_{i=0}^n \exp(\beta\cdot\gamma_i) \prod_{j=0(j\neq i)}^n (\gamma_i-\gamma_j)^{-1}. \tag{34}$$

All other cases are suitable limiting cases of this special case.

Proof: The proof of this lemma involves the use of the Laplace transform operator $\mathcal{L}$ defined for a function $f: \mathbb{R}_+ \to \mathbb{C}$ such that $\beta \to \exp(-\beta s)f(\beta)$ is in $L^1(0,\infty)$ for $s \in \mathbb{C}$ with Res sufficiently large:

$$(\mathcal{L}f)(s) = \int_0^\infty d\beta\, \exp(-\beta s)\cdot f(\beta).$$

Let $\theta(\beta) = 1$ for $\beta \ge 0$ and define the integration operator I as well as the multiplication operator M_α by $(If)(\beta) = \int_0^\beta d\beta' f(\beta')$ and $(M_\alpha f)(\beta) = \exp(\alpha\beta)\cdot f(\beta)$. If N and T_α are similarly defined through $(Ng)(s) = s^{-1}\cdot g(s)$ and $(T_\alpha g)(s) = g(s-\alpha)$, it is readily verified that $(\mathcal{L}\theta)(s) = s^{-1} = (N1)(s)$ and $\mathcal{L}\circ I = N\circ\mathcal{L}$ as well as $\mathcal{L}\circ M_\alpha = T_\alpha\circ\mathcal{L}$.

In terms of this notation,

$$\begin{aligned} &\mathcal{T}_n(\beta;\alpha_1\cdots\alpha_n) \\ &\quad = \int_0^\beta d\beta_1 \exp(\alpha_1\beta_1)\cdot\int_0^{\beta_1} d\beta_2 \exp(\alpha_2\beta_2)\cdots \\ &\quad\quad \times \int_0^{\beta_{n-1}} d\beta_n \exp(\alpha_n\beta_n)\cdot\theta(\beta_n) \end{aligned}$$

$$
\begin{aligned}
&= (I\circ M_{\alpha_1})\circ(I\circ M_{\alpha_2})\circ\cdots\circ(I\circ M_{\alpha_n})\,\theta(\beta)\\
&= \mathcal{L}^{-1}\circ\mathcal{L}\circ I\circ M_{\alpha_1}\circ\cdots\circ I\circ M_{\alpha_n}\theta(\beta)\\
&= \mathcal{L}^{-1}\circ N\circ T_{\alpha_1}\circ\cdots\circ N\circ T_{\alpha_n}\circ N1(\beta)\\
&= (2\pi i)^{-1}\int_{s_0-i_\infty}^{s_0+i_\alpha} dz\,\exp(z\beta)\cdot\prod_{i=0}^{n}(z-\gamma_i)^{-1}\\
&= \operatorname*{Res}_{s}\left(s\to\exp(s\beta)\cdot\prod_{i=0}^{n}(s-\gamma_i)^{-1}\right).
\end{aligned}
$$

The next to last step results from an application of the Mellin-inversion formula for $\mathcal{L}^{-1}$ with $s_0 > \max_i(\mathrm{Re}\gamma_i)$ to the function

$$
\begin{aligned}
N\circ T_{\alpha_1}\circ\cdots\circ N\circ T_{\alpha_n}\circ N1(s)
&= s^{-1}\cdot(s-\alpha_1)^{-1}\cdot\cdots\cdot(s-\alpha_1-\cdots-\alpha_n)^{-1}\\
&= \prod_{i=0}^{n}(s-\gamma_i)^{-1}.
\end{aligned}
$$

The last step involves a shift of the integration contour in the sense of $s_0\to-\infty$ and results from the Cauchy integral theorem.

Clearly, the special case of all γ_i distinct leads to the explicit evaluation of the residue given in Eq. (34). ///

This lemma is to be applied to Eq. (33) with $\alpha_i = (\tau_{i-1}-\tau_i)\cdot\epsilon$ or $\gamma_j = (\tau_0-\tau_j)\cdot\epsilon$. If Eq. (34) is used with the *proviso* that for coincident poles a suitable limiting case of the formula is meant, Eq. (33) leads to

$$
\begin{aligned}
&\int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n\,\tilde{E}_{T_\beta}\,(M_{\sigma_0\tau_0}\,\bar{P}(\beta_1)\cdots\bar{P}(\beta_n))\\
&\quad=\sum_{\tau_1\cdots\tau_n}\delta_{\sigma_0\tau_n}\,\bar{n}_{\sigma_0}(\beta)\cdot\bar{P}_{\tau_0\tau_1}\cdot\cdots\cdot\bar{P}_{\tau_{n-1}\tau_n}\\
&\quad\cdot\sum_{i=0}^{n}\exp[\beta(\tau_0-\tau_i)\cdot\epsilon]\times\prod_{\substack{j=0\\ j\neq i}}^{n}[(\tau_j-\tau_i)\cdot\epsilon]^{-1}.
\end{aligned}\tag{35}
$$

D. Summary

The single particle Hamiltonian h on $\mathcal{T}$ is assumed to generate a Gibbs semigroup $\mathbb{R}_+\ni\beta\to\exp(-\beta h)$ on $\mathcal{T}$. In terms of the eigenvector basis $(e_k)_{k\in K}$ of $\mathcal{T}$, $he_k = \epsilon_k e_k$.

The second quantized unperturbed Hamiltonian $H = d\bar{\Lambda}h$ on Fock space $\bar{\Lambda}\mathcal{T}$ then also generates a Gibbs semigroup, the unperturbed Gibbs semigroup $\beta\to S(\beta) = \exp(-\beta H)$. The unperturbed grand partition function is defined as $Z(\beta) = \mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[\exp(-\beta H)]$. The unperturbed Gibbs state $\tilde{E}_{T_\beta}$ on the Fock algebra $\mathfrak{F}(\mathcal{T})$ is defined in Eq. (10′) as

$$
\tilde{E}_{T_\beta}(A) = \frac{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}[\exp(-\beta H)A]}{Z(\beta)}.
$$

Let the perturbation $\bar{P} = -P$ be given as

$$
\bar{P} = \sum_{\sigma,\tau\in\mathbb{Z}_2^N}\bar{P}_{\sigma\tau}M_{\sigma\tau}
$$

with $\bar{P}_{\sigma\tau} = \bar{P}^*_{\tau\sigma}$ in terms of the matrix units $M_{\sigma\tau}$ for $\mathfrak{F}_N$ constructed in Lemma 3.2. Such a perturbation clearly satisfies all the hypotheses of Theorem 3.1(b). Then $\beta\to S^{\bar{P}}(\beta) = \exp[-\beta(H+P)]$ is the perturbed Gibbs semigroup with the perturbation expansion given in Eqs. (13) and (14).

The reduction Eq. (23) of the pertubed Gibbs state to the unperturbed Gibbs state is valid and the individual series terms can be calculated by means of formula (35).

When $\tau\in(\mathbb{Z}_2)^N$,

$$
\tau\cdot\epsilon = \sum_{k\in K_N}\tau_k\cdot\epsilon_k\quad\text{and}\quad\bar{n}_\tau(\beta) = \tilde{E}_{T_\beta}(M_{\tau\tau})
$$

is given by Eq. (32′). With $\dot{\Sigma}'$ indicating Σ subject to the proviso discussed in the paragraph following the proof of Lemma 3.4 we have

Theorem 3.2: Let H, P, etc., be as discussed above. Then

(i) $\mathbb{R}_+\ni\beta\to\exp[-\beta(H+P)]$ has the absolutely trace-norm convergent expansion

$$
\begin{aligned}
\exp[-\beta(H+P)] = \sum_{n=0}^{\infty}\int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n\\
\times S(\beta-\beta_1)\bar{P}S(\beta_1-\beta_2)\bar{P}\cdots\bar{P}S(\beta_n),
\end{aligned}\tag{36}
$$

where $\Delta_n^\beta = \{(\beta_1,\ldots,\beta_n)\,|\,\beta\geq\beta_1\geq\beta_2\geq\cdots\geq\beta_n\geq 0\}$.

(ii) For $A\in\mathfrak{F}(\mathcal{T})$, the Fock algebra of $\mathcal{T}$, with absoutely convergent sums

$$
\begin{aligned}
&\frac{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}\{\exp[-\beta(H+P)]A\}}{\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}\{\exp[-\beta(H+P)]\}}\\
&\quad=\frac{\sum_{n=0}^{\infty}\int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n\tilde{E}_{T_\beta}\,(\bar{P}(\beta_1)\cdots\bar{P}(\beta_n)A)}{\sum_{n=0}^{\infty}\int_{\Delta_n^\beta} d\beta_1\cdots d\beta_n\tilde{E}_{T_\beta}\,(\bar{P}(\beta_1)\cdots\bar{P}(\beta_n))}
\end{aligned}\tag{36′}
$$

is the reduction of the perturbed Gibbs state to the unperturbed Gibbs state E_{T_β}.

(iii) If $A = \sum_{\sigma_0\tau_0}A_{\sigma_0\tau_0}M_{\sigma_0\tau_0}\in\mathfrak{F}_N$ then with absolute convergence uniformly in A,

$$
\begin{aligned}
&\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}\{\exp[-\beta(H+P)]A\}\\
&\quad= Z(\beta)\cdot\sum_{n=0}^{\infty}\sum_{\tau_0\cdots\tau_n}\bar{P}_{\tau_0\tau_1}\cdots\bar{P}_{\tau_{n-1}\tau_n}\cdot\bar{n}_{\tau_n}(\beta)\cdot A_{\tau_n\tau_0}\\
&\quad\times{\sum}'\frac{\exp[\beta\epsilon\cdot(\tau_0-\tau_i)]}{\prod_{\substack{j=1\\ j\neq i}}^{n}(\tau_j-\tau_i)\cdot\epsilon}.
\end{aligned}\tag{36″}
$$

4. CONCLUSION

Formula (36″) provides a convergent resolution of the (unnormalized) correlation function $\mathrm{Tr}_{\bar{\Lambda}\mathcal{T}}\{\exp[-\beta(H+P)]\cdot A\}$ into simpler terms which are explicitly calculable once $P_{\tau\tau'}$ and $A_{\tau\tau'}$ are specified. The thermodynamic information is coded into $\bar{n}_\tau(\beta)$ and the

$$
{\sum}'\frac{\exp[\beta\epsilon\cdot(\tau_0-\tau_i)]}{\prod_{\substack{j=1\\ j\neq i}}^{n}(\tau_j-\tau_i)\cdot\epsilon}
$$

factors. Diagram rules could easily be formulated so as to mechanize the computation of any given term. The present treatment differs from most attempts in the same direction in that a convergence proof is furnished in a mathematically precise fashion.

Generalizations to more complicated interactions appear to be within reach. Likewise an analogous treatment of the boson situation will be given subsequently.

* Supported in part by Grant GP 18641 of the National Science Foundation.

1 D. A. Uhlenbrock, J. Math. Phys. 12, 2503 (1971).
2 D. Shale and W. F. Stinespring, Ann. Math. 80, 365 (1964).
3 R. T. Powers, "Representations of the Canonical Anticommutation Relations," Thesis, Dept. of Physics, Princeton University (June, 1967).
4 G. F. Dell'Antonio, Commun. Math. Phys. 9, 81 (1968); G. Rideau, Commun. Math. Phys. 9, 229 (1968).
5 E. Balslev, J. Manuceau, and A. Verbeure, Commun. Math. Phys. 8, 315 (1968).
6 R. T. Powers and E. Størmer, "Free States of the Canonical Anticommutation Relations," U. of Pennsylvania, preprint (1970).
7 E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* Revised edition (Amer. Math. Soc. Colloquium, Providence, R.I., 1957), Vol. 31.
8 D. Kastler, J. C. T. Pool, and E. T. Poulsen, Commun. Math. Phys. 12, 175 (1969). The precursor to this paper is R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. 5, 215 (1967).

The Plasma Inverse Problem

George Nick Balanis*
Applied Theory, Inc., Westwood, Los Angeles, California
and
California Institute of Technology, Pasadena, California

(Received 14 February 1972)

A simple and direct development of the theory of the plasma inverse problem is given, and it is shown that the time record of the reflected wave arising from a δ function electric field normally incident on a stratified plasma determines uniquely the plasma density through an integral equation.

I. INTRODUCTION

Many interesting physical problems involve the propagation of waves in a medium, and we usually solve these problems to find the propagating fields. However, for some problems we can deduce properties of the medium from a limited knowledge of the propagating fields. These problems where we know properties of the fields and want to find properties of the medium are called inverse problems, and there is a variety of them.

Lord Rayleigh[1] and Krein[2] determine the density of an inhomogeneous vibrating string from the string's eigenfrequencies while Ambarzumian,[3] Borg,[4] and Levitan and Gasymov[5] show how to obtain the potential in Schrödinger's equation from two sets of eigenvalues. Eaves, Twardeck, and Morin[6] consider a plasma in a cavity and determine the plasma density from the cavity resonant frequencies. From the surface potential resulting from an electrode supplying direct current to the earth's surface, Langer[7,8] shows how to find the conductivity of the earth as a function of depth. Gel'fand and Levitan[9] and Marchenko[10,11] describe a procedure for obtaining Schrödinger's potential from the wavefunction form at large distances. By using the impulse response current of an inhomogeneous transmission line, Gopinath and Shondhi[12] determine the line capacitance.

In 1951, Kay[13,14] considers the possibility of determining the electron density of a plane stratified plasma from the reflected wave that results when a δ function electric field is normally incident on the plasma. Under certain approximations the plasma electric field $E(z,t)$ generated by the incident δ function obeys the plasma wave equation

$$\frac{\partial^2 E(z,t)}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E(z,t)}{\partial t^2} - V(z)E(z,t) = 0,$$

where z is the coordinate along the axis of stratification, c the speed of light in air, and $V(z)$ a function directly related to the plasma density. Kay is able to show that if the reflected wave $R(ct)$ is known as a function of time, then

$$V(z) = 2\frac{d}{dz}K(z,z),$$

where $K(z,y)$ is the solution of the integral equation

$$\text{(i)}\; R(z+y) + K(z,y) + \int_{-z}^{z} K(z,z')R(z'+y)\,dz' = 0, \qquad -z < y < z.$$

For certain special reflected waves corresponding to the case when the reflection coefficient of the plasma is a rational function,[15] Kay is able to obtain an analytical solution of the integral equation, and thus obtain $V(z)$ and the plasma density.

Balanis[16] develops a general solution of the integral equation (i) for plane stratified plasmas. The solution is found by converting the integral equation to a matrix equation through an appropriate numerical discretization. The matrix equation, which relates the values of the reflected wave at equally spaced time points to discrete values of the function K, is inverted and thus the functions K and $V(z)$ are obtained. Such a solution is essential because even for the simplest $V(z)$, the plasma reflection coefficient is not a rational function of frequency; thus Kay's analytical solution is not applicable, and because a graph of the time record of the reflected wave will be available from which the time sampled values of the reflected wave can be easily obtained. Also Balanis examines the case when the incident wave is a thin square pulse, instead of a δ function.

This paper has the limited objective of reconsidering with another approach the plasma inverse problem examined by Kay. The approach developed in this paper is simple and direct and yet powerful enough to illuminate all the essential features of continuous inverse problems. (It is worth noting that even in the basic paper of Gel'fand[9] the theory is developed with an indirect approach.)

II. ANALYSIS

Let a collisionless, unbiased and with zero initial electric and magnetic fields, plane stratified plasma of electron density $N(z)(el/m^3)$[17] occupy the region $z \geq 0$. The space $z < 0$ as well as the space between the charges are filled with air. Let a δ function electric field

$$E_{\text{inc}}(z, ct) = \delta(z - ct)$$

* Supported in part by Grant GP 18641 of the National Science Foundation.

1 D. A. Uhlenbrock, J. Math. Phys. **12**, 2503 (1971).
2 D. Shale and W. F. Stinespring, Ann. Math. **80**, 365 (1964).
3 R. T. Powers, "Representations of the Canonical Anticommutation Relations," Thesis, Dept. of Physics, Princeton University (June, 1967).
4 G. F. Dell'Antonio, Commun. Math. Phys. **9**, 81 (1968); G. Rideau, Commun. Math. Phys. **9**, 229 (1968).
5 E. Balslev, J. Manuceau, and A. Verbeure, Commun. Math. Phys. **8**, 315 (1968).
6 R. T. Powers and E. Størmer, "Free States of the Canonical Anticommutation Relations," U. of Pennsylvania, preprint (1970).
7 E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* Revised edition (Amer. Math. Soc. Colloquium, Providence, R.I., 1957), Vol. 31.
8 D. Kastler, J. C. T. Pool, and E. T. Poulsen, Commun. Math. Phys. **12**, 175 (1969). The precursor to this paper is R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).

The Plasma Inverse Problem

George Nick Balanis*
Applied Theory, Inc., Westwood, Los Angeles, California
and
California Institute of Technology, Pasadena, California
(Received 14 February 1972)

A simple and direct development of the theory of the plasma inverse problem is given, and it is shown that the time record of the reflected wave arising from a δ function electric field normally incident on a stratified plasma determines uniquely the plasma density through an integral equation.

I. INTRODUCTION

Many interesting physical problems involve the propagation of waves in a medium, and we usually solve these problems to find the propagating fields. However, for some problems we can deduce properties of the medium from a limited knowledge of the propagating fields. These problems where we know properties of the fields and want to find properties of the medium are called inverse problems, and there is a variety of them.

Lord Rayleigh[1] and Krein[2] determine the density of an inhomogeneous vibrating string from the string's eigenfrequencies while Ambarzumian,[3] Borg,[4] and Levitan and Gasymov[5] show how to obtain the potential in Schrödinger's equation from two sets of eigenvalues. Eaves, Twardeck, and Morin[6] consider a plasma in a cavity and determine the plasma density from the cavity resonant frequencies. From the surface potential resulting from an electrode supplying direct current to the earth's surface, Langer[7,8] shows how to find the conductivity of the earth as a function of depth. Gel'fand and Levitan[9] and Marchenko[10,11] describe a procedure for obtaining Schrödinger's potential from the wavefunction form at large distances. By using the impulse response current of an inhomogeneous transmission line, Gopinath and Shondhi[12] determine the line capacitance.

In 1951, Kay[13,14] considers the possibility of determining the electron density of a plane stratified plasma from the reflected wave that results when a δ function electric field is normally incident on the plasma. Under certain approximations the plasma electric field $E(z,t)$ generated by the incident δ function obeys the plasma wave equation

$$\frac{\partial^2 E(z,t)}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E(z,t)}{\partial t^2} - V(z)E(z,t) = 0,$$

where z is the coordinate along the axis of stratification, c the speed of light in air, and $V(z)$ a function directly related to the plasma density. Kay is able to show that if the reflected wave $R(ct)$ is known as a function of time, then

$$V(z) = 2\frac{d}{dz}K(z,z),$$

where $K(z,y)$ is the solution of the integral equation

$$\text{(i)}\; R(z+y) + K(z,y) + \int_{-z}^{z} K(z,z')R(z'+y)\,dz' = 0, \qquad -z < y < z.$$

For certain special reflected waves corresponding to the case when the reflection coefficient of the plasma is a rational function,[15] Kay is able to obtain an analytical solution of the integral equation, and thus obtain $V(z)$ and the plasma density.

Balanis[16] develops a general solution of the integral equation (i) for plane stratified plasmas. The solution is found by converting the integral equation to a matrix equation through an appropriate numerical discretization. The matrix equation, which relates the values of the reflected wave at equally spaced time points to discrete values of the function K, is inverted and thus the functions K and $V(z)$ are obtained. Such a solution is essential because even for the simplest $V(z)$, the plasma reflection coefficient is not a rational function of frequency; thus Kay's analytical solution is not applicable, and because a graph of the time record of the reflected wave will be available from which the time sampled values of the reflected wave can be easily obtained. Also Balanis examines the case when the incident wave is a thin square pulse, instead of a δ function.

This paper has the limited objective of reconsidering with another approach the plasma inverse problem examined by Kay. The approach developed in this paper is simple and direct and yet powerful enough to illuminate all the essential features of continuous inverse problems. (It is worth noting that even in the basic paper of Gel'fand[9] the theory is developed with an indirect approach.)

II. ANALYSIS

Let a collisionless, unbiased and with zero initial electric and magnetic fields, plane stratified plasma of electron density $N(z)(el/m^3)$[17] occupy the region $z \geq 0$. The space $z < 0$ as well as the space between the charges are filled with air. Let a δ function electric field

$$E_{\text{inc}}(z, ct) = \delta(z - ct)$$

polarized normally to the z axis be incident on the plasma from the region $z < 0$. An electric field E_{refl} polarized in the same direction as the incident field is reflected and propagates towards the left:

$$E_{\text{refl}}(z, ct) = R(z + ct).$$

We show that the time record of the reflected wave at $z = 0$ determines (uniquely) the electron density $N(z)$ at every finite z inside the plasma.

The incident δ function generates a plasma electric field which propagates towards the right, i.e., towards increasing z. The shape of the generated plasma electric field is governed by Maxwell's equations applied inside the plasma, the continuity of the electric and magnetic fields across the air-plasma interface at $z = 0$, and the initial condition that at $t = 0$ the plasma field is zero. A simple application of Maxwell's equations shows that the plasma electric field is obtained by solving the plasma wave equation (1), the continuity equations (2) and (3), and the initial condition (4):

$$\frac{\partial^2 E(z,ct)}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E(z,ct)}{\partial t^2} - k_p^2(z) E(z,ct) = 0, \quad z \geq 0, \tag{1}$$

$$E(0, ct) = \delta(-ct) + R(ct), \tag{2}$$

$$\frac{\partial E(0,ct)}{\partial z} = \delta'(-ct) + R'(ct), \tag{3}$$

$$E(z, 0) = 0, \quad z > 0, \tag{4}$$

where $k_p^2(z)$ is assumed to be a nonnegative, bounded, continuous, and with continuous derivative function of z directly related to the electron density $N(z)$,

$$k_p^2(z) = \frac{q^2 \mu_0}{m_e} N(z) \tag{5}$$

with

q = electron charge (coul),

μ_0 = permeability of free air (henry/met),

m_e = electron mass (kgr).

Equation (1) is a linear hyperbolic partial differential equation of the second order, and it is well known that for this equation the initial condition (4) can be replaced without any error by the causality condition (4′)

$$E(z, ct) = 0, \quad t < z/c, \quad z > 0. \tag{4'}$$

Before we proceed further with the inverse problem we need to show several interesting aspects of the plasma wave equation (1). We start by proving the following theorem.

Theorem 1: Let $E_1(z, ct)$ be the solution of the plasma wave equation with boundary conditions (7) and (8) and $E_2(z, ct)$ the solution with boundary conditions (10) and (11):

$$\frac{\partial^2 E_1}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E_1}{\partial t^2} - k_p^2 E_1 = 0, \quad z \geq 0, \tag{6}$$

$$E_1(o, ct) = \delta(-ct), \tag{7}$$

$$\frac{\partial E_1(o,ct)}{\partial z} = \delta'(-ct), \tag{8}$$

and

$$\frac{\partial_2 E_2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E_2}{\partial t^2} - k_p^2 E_2 = 0, \quad z \geq 0, \tag{9}$$

$$E_2(o, ct) = \delta(ct), \tag{10}$$

$$\frac{\partial E_2(o,ct)}{\partial z} = \delta'(ct). \tag{11}$$

Then any time-dependent solution $E(z, ct)$ of the plasma wave equation (12)

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} - k_p^2 E = 0, \quad z \geq 0, \tag{12}$$

is obtained as a linear combination of $E_1(z, ct)$, $E_2(z, ct)$ from (13):

$$E(z, ct) = \int_{-\infty}^{+\infty} A(ct - y) E_1(z, y) dy + \int_{-\infty}^{+\infty} B(ct - y) \times E_2(z, y) dy, \tag{13}$$

where $A(ct)$, $B(ct)$ are arbitrary functions of time and independent of z.

Proof: Let $\hat{E}_1(z,k), \hat{E}_2(z,k), \hat{E}(z,k)$ be the Fourier transforms of $E_1(z, ct), E_2(z, ct)$, and $E(z, ct)$, respectively, where

$$\hat{E}_1(z, k) = \int_{-\infty}^{+\infty} E_1(z, ct) e^{ikct} c\, dt,$$

$$\hat{E}_2(z, k) = \int_{-\infty}^{+\infty} E_2(z, ct) e^{ikct} c\, dt,$$

$$\hat{E}(z, k) = \int_{-\infty}^{+\infty} E(z, ct) e^{ikct} c\, dt,$$

and $k = \omega/c$.

By taking the Fourier transform of Eqs. (6)–(12), we find

$$\frac{\partial^2 \hat{E}_1}{\partial z^2} + [k^2 - k_p^2]\hat{E}_1 = 0, \quad z \geq 0, \tag{14}$$

$$\hat{E}_1(o, k) = 1, \tag{15}$$

$$\frac{\partial \hat{E}_1(o,k)}{\partial z} = ik \tag{16}$$

and

$$\frac{\partial^2 \hat{E}_2}{\partial z^2} + [k^2 - k_p^2]\hat{E}_2 = 0, \quad z \geq 0, \tag{17}$$

$$\hat{E}_2(o, k) = 1, \tag{18}$$

$$\frac{\partial \hat{E}_2}{\partial z}(o, k) = -ik, \tag{19}$$

and

$$\frac{\partial^2 \hat{E}}{\partial z^2} + [k^2 - k_p^2]\hat{E} = 0, \quad z \geq 0. \tag{20}$$

Let $W(z, k)$ be the Wronskian of the solutions E_1, E_2, where

$$W(z, k) \equiv E_2 \frac{\partial E_1}{\partial z} - E_1 \frac{\partial E_2}{\partial z}. \tag{21}$$

By differentiating with respect to z, Eq. (21) and with the help of (14) and (17), we obtain

$$\frac{\partial W}{\partial z}(z,k) = 0,$$

and thus $W(z,k)$ is a constant independent of z. By evaluating $W(z,k)$ at $z = 0$ with the help of Eqs.(15), (16), (18), and (19) we obtain

$$W(z,k) = 2ik. \tag{22}$$

It is well known that if the Wronskian of two solutions of an ordinary differential equation is not equal to zero, then the solutions form a fundamental set, and any other solution is obtained by a linear combination of these fundamental solutions. Thus the solutions $\hat{E}_1, \hat{E}_2$ form a fundamental set of (20) for $k \neq 0$, and we write the solution $\hat{E}$ as

$$\hat{E}(z,k) = a(k)\hat{E}_1(z,k) + b(k)\hat{E}_2(z,k) + 2\pi f(z)\delta(k), \tag{23}$$

where $a(k), b(k)$ are arbitrary functions of k and $f(z)$ an arbitrary function of z introduced to account for the fact that the Wronskian vanishes at $k = 0$. By taking the inverse Fourier transform of (23) and neglecting the time-independent term $f(z)$, we obtain (13), and this completes the proof.

As can be easily seen from Eqs.(14)–(19) and Eq.(5) the solutions $\hat{E}_1, \hat{E}_2$ depend only on the plasma density. This dependence is made more explicit if we note that a plasma behaves like a high pass filter, letting the high frequencies of any wave inside it propagate without any attenuation,[18] and we write

$$\hat{E}_1(z,k) = e^{ikz} + \hat{C}_1(z,k), \tag{24}$$

and

$$\hat{E}_2(z,k) = e^{-ikz} + \hat{C}_2(z,k) \tag{25}$$

with $\hat{C}_1(z,k)$ and $\hat{C}_2(z,k)$ having no appreciable high frequencies. The terms $\hat{C}_1, \hat{C}_2$ in Eqs.(24) and (25) contain all the effects that the plasma contributes to the behavior of the solutions $\hat{E}_1$ and $\hat{E}_2$. Indeed if in the space $z \geq 0$ there is no plasma, then $k_p^2(z) = 0$, and from Eqs.(14)–(19) we obtain that $\hat{E}_1(z,k) = e^{+ikz}$ and $\hat{E}_2(z,k) = e^{-ikz}$ and therefore from (24) and (25) $\hat{C}_1(z,k) = \hat{C}_2(z,k) = 0$.

By taking the inverse Fourier transform of Eqs.(24) and (25), we find

$$E_1(z,ct) = \delta(z - ct) + C_1(z,ct), \tag{26}$$

$$E_2(z,ct) = \delta(z + ct) + C_2(z,ct), \tag{27}$$

and after we substitute the above equations into Eqs. (6)–(11) we find

$$\frac{\partial^2 C_1}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 C_1}{\partial t^2} - k_p^2 C_1 = k_p^2 \delta(z - ct), \tag{28}$$

$$C_1(0,ct) = 0, \tag{29}$$

$$\frac{\partial C_1}{\partial z}(0,ct) = 0 \tag{30}$$

and

$$\frac{\partial^2 C_2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 C_2}{\partial t^2} - k_p^2 C_2 = k_p^2 \delta(z + ct), \tag{31}$$

$$C_2(0,ct= 0, \tag{32}$$

$$\frac{\partial C_2}{\partial z}(0,ct) = 0. \tag{33}$$

The functions $C_1(z,ct)$ and $C_2(z,ct)$ are important in inverse problems. They will be referred to herein as characteristic fields because these functions characterize the plasma completely and because the plasma density is most easily obtained from the values of these functions along the characteristic lines of the plasma wave equation. We begin the discussion about the characteristic fields by listing several of their properties and, subsequently, proving them.

Property 1: The plasma density determines uniquely the characteristic fields.

Property 2: The plasma density is obtained uniquely from the values of the characteristic fields along the characteristic lines of the plasma wave equation.

Property 3: The characteristic fields are related to each other.

Proof of Property 1: We prove Property 1 for the field $C_1(z,ct)$. Similar arguments prove this property for $C_2(z,ct)$.

A central result of the theory of characteristics of hyperbolic partial differential equations states that the value of the solution of a second order partial differential equation at a point on some boundary line influences the behavior of the solution only in the region included between the characteristic curves of the partial differential equation emanating from that point. Now, the characteristics of the plasma wave equation are straight lines parallel to the lines $z = \pm ct$, and thus by applying the theory of characteristics to the solution $E_1(z,ct)$ of Eqs.(6)–(8) and identifying the boundary line as the ct-axis, we immediately obtain that

$$E_1(z,ct) = 0, \quad |ct| < z,$$

and thus from (26) we find

$$C_1(z,ct) = 0, \quad |ct| < z. \tag{34}$$

The δ function on the right-hand side of Eq.(28) indicates that except on the line $z = ct$, the field $C_1(z,ct)$ obeys the plasma wave equation; and, furthermore, that across the line $z = ct$, $C_1(z,ct)$ is discontinuous whereas across the line $z = -ct$ there is no discontinuity. Thus, we find that

$$\frac{\partial^2 C_1}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 C_1}{\partial t^2} - k_p^2 C_1 = 0, \quad -z < ct < z, \tag{35}$$

and

$$C_1(z,-z) = 0. \tag{36}$$

The jump discontinuity of $C_1(z,ct)$ across the line $z = ct$ is easily obtained if we let

$$C_1(z,ct) = H(z - ct)g(z,ct), \tag{37}$$

where $g(z,ct)$ is continuous solution of the plasma wave equation in the region $-z < ct < z$ and substitute (37) into (28).[19] After some simple algebra we find that

$$\frac{dg(z,z)}{dz} = \frac{1}{2}k_p^2(z)$$

and thus

$$\frac{dC_1(z,z)}{dz} = \frac{1}{2}k_p^2(z). \quad (38)$$

Now, grouping all the results together, we have that in the region $|ct| \le z$, the characteristic field $C_1(z,ct)$ is the solution of Eqs. (39)-(41),

$$\frac{\partial^2 C_1}{\partial z^2} - \frac{1}{C^2}\frac{\partial^2 C_1}{\partial t^2} - k_p^2 C_1 = 0, \quad z \ge 0, \ |ct| \le z, \quad (39)$$

$$\frac{dC_1(z,z)}{dz} = \frac{1}{2}k_p^2(z), \quad z \ge 0, \quad (40)$$

$$C_1(z,-z) = 0, \quad z \ge 0, \quad (41)$$

whereas outside this region, $C_1(z,ct)$ obeys Eq. (42):

$$C_1(z,ct) = 0, \quad z \ge 0, \ |ct| > z. \quad (42)$$

It turns out that Eqs. (39)–(41) constitute a Goursat problem for the characteristic field $C_1(z,ct)$, and it is known that such a problem has a unique, bounded, and continuous solution in any finite z interval. [Replace (40) by $C_1(z,z) = \frac{1}{2}\int_0^z k_p^2(\xi)\,d\xi$ and see Ref. 22.] Thus, for any finite z, the plasma density determines a unique, continuous, and bounded characteristic field $C_1(z,ct)$, and once $k_p^2(z)$ is given $C_1(z,ct)$ can be found by solving the Goursat problem (39)–(41).

Similar arguments show that inside the region $|ct| \le z$, $C_2(z,ct)$ is the solution of the Goursat problem (43)–(45) and zero outside:

$$\frac{\partial^2 C_2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 C_2}{\partial t^2} - k_p^2 C_2 = 0, \quad z \ge 0, \ |ct| \le z, \quad (43)$$

$$\frac{d}{dz}C_2(z,-z) = \frac{1}{2}k_p^2(z), \quad z \ge 0, \quad (44)$$

$$C_2(z,z) = 0, \quad z \ge 0, \quad (45)$$

$$C_2(z,ct) = 0, \quad z \ge 0, \ |ct| > z; \quad (46)$$

and thus the proof of Property 1 is completed.

Proof of Property 2: From Eqs. (40) and (44) we see that if the characteristic field $C_1(z,ct)$ is known along the line $z = ct$ and $C_2(z,ct)$ along the line $z = -ct$, then a simple differentiation along these lines yields $k_p^2(z)$, and thus we obtain uniquely $N(z)$ through (5)[20] and the proof of Property 2 is complete.

Proof of Property 3: Replacing t by $-t$ in Eqs. (39)–(42) for the characteristic field $C_1(z,ct)$ we obtain Eqs. (43)–(46) for the characteristic field $C_2(z,ct)$. Thus the characteristic fields are related by the simple time inversion (47):

$$C_2(z,ct) = C_1(z,-ct) \quad (47)$$

and hence Property 3 is shown.[21]

Now we return to the inverse scattering problem. By applying Theorem 1 to the plasma electric field, solving Eqs. (1)–(5), we obtain

$$E(z,ct) = \dot{E}_1(z,ct) + \int_{-\infty}^{+\infty} R(ct-y)E_2(z,y)dy,$$

and with the help of Eqs. (26), (27), and (47) the above equation becomes

$$E(z,ct) = \delta(z-ct) + C_1(z,ct) + \int_{-\infty}^{+\infty} C_1(z,y)R(y+ct)dy + R(z+ct).$$

Via Eq. (42) we finally obtain that the plasma electric field is given by (48),

$$E(z,ct) = \delta(z-ct) + C_1(z,ct) + R(z+ct) + \int_{-z}^{z} C_1(z,y) + R(y+ct)dy. \quad (48)$$

By restricting the plasma field to a right traveling wave (i.e., a wave moving toward $z = \infty$) by the causality condition (4′) we obtain

$$C_1(z,ct) + R(z+ct) + \int_{-z}^{z} C_1(z,y)R(y+ct)dy = 0, \quad ct < z. \quad (49)$$

We remarked before that the high frequencies of the δ function incident electric field propagate through the plasma, and thus the reflected wave $R(ct)$ does not possess appreciable high frequencies; and so it is reasonable to make the statement that the reflected wave is a continuous function of time in any finite time interval. In fact, in Ref. 16 it is proved that the reflected wave $R(ct)$ is a continuous function of time when $N(z)$ is a nonnegative bounded, piecewise continuous function of z which tends toward zero for large z; and since from physical arguments it is evident that $R(ct)$ cannot depend on the values of the electron density for $z > ct/2$, a simple extension of the proof given in Ref. 16 shows that the reflected wave $R(ct)$ is continuous function of time in any finite time interval even in the case when the electron density does not tend to zero for large z. So, by applying the causality condition to the electric field at $z = 0$, we obtain

$$R(ct) = 0, \quad ct \le 0, \quad (50)$$

and incorporating (50) into (49) and using (42) we find

$$C_1(z,ct) + R(z+ct) + \int_{-ct}^{z} C_1(z,y)R(y+ct)dy = 0, \quad -z \le ct < z.$$

In the above equation the functions $C_1(z,ct)$ and $R(ct)$ are continuous functions of their arguments, and thus it appears reasonable to state that the above integral equation holds also for $ct = z$. In fact, this is proved in Ref. 16. So, we finally obtain the integral Eq. (51):

$$C_1(z,ct) + R(z+ct) + \int_{-ct}^{z} C_1(z,y)R(y+ct)dy = 0, \quad -z \le ct \le z \quad (51)$$

Equation (51) is a Fredholm integral equation of the second kind, and in Ref. 12 it is shown that it has a unique continuous solution when $N(z)$ is a nonnegative, piecewise, continuous function of z that tends to zero sufficiently fast as $z \to 00$. However, since both $R(ct)$ for $0 \le ct \le 2z_0$ and $C_1(z_0,ct)$ for $-z_0 \le ct \le z_0$ cannot depend on the values of the electron density for $z > z_0$, a simple extension of the proof found in Ref. 12 shows that (51) has a unique continuous solution $C_1(z,ct)$ for any finite z when $N(z)$ is a nonnegative bounded, continuous, and with continuous derivative function of z. Since we have already constructed one solution, namely the characteristic field $C_1(z,ct)$ of the plasma wave equation, then this characteristic field is the only solution.

Grouping all our results together we state that given the time record of the reflected wave $R(ct)$ for $0 \le t \le t_0$ of an electron density $N(z)$ which is zero for $z < 0$, then $N(z)$ is uniquely obtained for all $z, z < \frac{1}{2} ct_0$, by solving (52)–(54):

$$R(z + ct) + C_1(z, ct) + \int_{-ct}^{z} C_1 z, z')R(z' + ct)dz' = 0, \quad - z \le ct \le z,\ z \ge 0; \tag{52}$$

$$\frac{d}{dz} C_1(z, z) = \tfrac{1}{2} k_p^2(z); \tag{53}$$

$$k_p^2(z) = \frac{q^2 \mu_0}{m_e} N(z); \tag{54}$$

where $R(ct)$ is a continuous function of time for any finite time interval and $C_1(z, ct)$ a continuous function of z and ct for any finite z interval and $-z \le ct \le +z$.

The results stated in the previous paragraph are essentially unchanged if one assumes that the function $N(z)$ is allowed to be more general, that is a nonnegative, bounded, and piecewise continuous function of z.[12,16]

ACKNOWLEDGMENTS

The author expresses his thanks for helpful suggestions of his advisor Professor C. H. Papas and Professor G. Gavalas and Professor G. B. Whitham of the California Institute of Technology. This research was supported by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR-70-1935 to the California Institute of Technology.

* The author is a Research Fellow at the California Institute of Technology and a member of the technical staff of Applied Theory, Inc., Westwood, Los Angeles, California.

1 Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. I, 214–16.
2 M. G. Krein, Dokl. Akad. Nauk SSSR **76**, 345 (1951).
3 V. A. Ambarzumian, Z. Phyzik **53**, 690 (1929).
4 G. Borg, Acta Math. **78**, 1 (1946).
5 B. M. Levitan and M. G. Gasymov, Russ Math. Surv. **19**, No. 2, 1 (1964).
6 R. E. Eaves, T. G. Twardeck, and S. J. Morin, IEEE Trans. Antennas Propagation **19**, No. 2 (1971).
7 R. E. Langer, Amer. Math. Soc. Bull. **39**, 814 (1933).
8 L. B. Slichter, Physics **4**, 307 (1933).
9 I. M. Gel'fand and B. M. Levitan, Amer. Math. Soc. Transl., Ser. 2, No. 1, 253 (1955).
10 V. A. Marchenko, Tr. Mosk. Natural'n Obs. **1**, 327 (1952); *ibid.* **2**, 3 (1953).
11 L. D. Faddeyev, "The Inverse Problem in the Quantum Theory of Scattering," Research Report No. EM-165, NYU Div. of EM Research (1960).
12 B. Gopinath and M. Sondhi, Proc. IEEE **59**, No. 3, 383 (1971).
13 Irvin Kay, "The Inverse Scattering Problem," Research Report No. EM-74, NYU Div. of EM Research (1955).
14 Irvin Kay, "On the Determination of the Free Electron Distribution of an Ionized Gas," Research Report No. EM-141, NYU Div. of EM Research (1959).
15 Irvin Kay, Commun. Pure Appl. Math. **12**, 371 (1960).
16 G. N. Balanis, "Plasma Inverse Scattering Theory," PhD. Thesis in EE Dept. of California Institution of Tech. (June, 1972).

17 The electron density $N(z)$ is assumed to be a nonnegative, bounded, continuous, and with continuous derivative function of z throughout this paper.

18 It can be easily seen from (14)–(19) that $|C_1|, |C_2| \to 0$ as $k \to \pm\infty$.

19 $H(\xi)$ is the Heaviside step function,

$$H(\xi) = \begin{cases} 0, & \xi < 0 \\ 1, & \xi > 0 \end{cases}.$$

20 It is of interest to note that we could also obtain $k_p^2(z)$ from the characteristic fields by solving for $k_p^2(z)$ in Eqs. (39) and (43); however, such a procedure is rather cumbersome.

21 It is evident that this essential property of the characteristic fields is a consequence of the invariance of the plasma wave equation under a time reversal.

22 S. G. Mikhlin, *Linear Equations of Mathematical Physics* (Holt and Rinehart, New York, 1967), p. 27.

Lattice Wind-Tree Models. I. Absence of Diffusion

D. J. Gates*

Mathematics Department, Rockefeller University, New York, N.Y. 10021

(Received 19 January 1972; Revised Manuscript Received 6 March 1972)

Some lattice versions of the wind-tree model of Ehrenfest are introduced and studied. They include two versions in which overlapping of tree particles is forbidden, a third in which it is allowed, and a fourth in which it is allowed but results in a finite repulsive force. In every case it is found that the mean-square displacement $\Delta(t)$ of the wind particles at time t is bounded independently of t at sufficiently high density of the trees. This is in sharp contrast with the Einstein relation $\Delta(t) = O(t)$ as $t \to \infty$, which might be expected to hold at low densities. Randomization of the initial velocity of a wind particle is also shown to occur in a certain sense, and upper bounds on a recurrence time are obtained at high density. On the other hand, it is shown that thermalization does not occur even at low densities.

1. INTRODUCTION

The Lorentz model and the "wind-tree" model of Ehrenfest have been used recently for the study of the diffusion process and the problem of approach to equilibrium (e.g., Refs. 1, 2, 3). In these models there are two types of particles, called wind particles and tree particles (or trees). The trees are at rest and distributed in equilibrium under the action of their mutual forces and are unaffected by the motion of the wind particles. The latter have no mutual forces, so that their evolution is determined entirely by their interaction (or collisions) with the trees and the container walls.

We consider a system of square trees with sides of length 1, centered on the points (Z^2) of a simple square lattice with unit spacing. A typical configuration is illustrated in Fig. 1 where the lattice is shown inclined at 45°. We specify that the system be enclosed by a box consisting of a close-packed arrangement of trees; but the over-all shape of the box is otherwise arbitrary.

We consider a wind particle which sits initially at some point $\mathbf{q}$ and has initial velocity $\mathbf{p}$ along one of the four lattice directions at 45° to the lattice lines. The wind particle is deflected through a right-angle whenever it strikes a tree, as shown in Fig. 1. Its position $\mathbf{q}_t(\mathbf{x}, C, \Lambda)$ and velocity $\mathbf{p}_t(\mathbf{x}, C, \Lambda)$ at time t depend only on the initial dynamical state $\mathbf{x} \equiv (\mathbf{q}, \mathbf{p})$, the set C

the plasma wave equation, then this characteristic field is the only solution.

Grouping all our results together we state that given the time record of the reflected wave $R(ct)$ for $0 \le t \le t_0$ of an electron density $N(z)$ which is zero for $z < 0$, then $N(z)$ is uniquely obtained for all $z, z < \frac{1}{2} ct_0$, by solving (52)–(54):

$$R(z + ct) + C_1(z, ct) + \int_{-ct}^{z} C_1 z, z')R(z' + ct)dz' = 0,$$
$$- z \le ct \le z, \ z \ge 0; \qquad (52)$$
$$\frac{d}{dz} C_1(z, z) = \tfrac{1}{2} k_p^2(z); \qquad (53)$$
$$k_p^2(z) = \frac{q^2 \mu_0}{m_e} N(z); \qquad (54)$$

where $R(ct)$ is a continuous function of time for any finite time interval and $C_1(z, ct)$ a continuous function of z and ct for any finite z interval and $-z \le ct \le + z$.

The results stated in the previous paragraph are essentially unchanged if one assumes that the function $N(z)$ is allowed to be more general, that is a nonnegative, bounded, and piecewise continuous function of z.[12,16]

ACKNOWLEDGMENTS

The author expresses his thanks for helpful suggestions of his advisor Professor C.H. Papas and Professor G. Gavalas and Professor G.B. Whitham of the California Institute of Technology. This research was supported by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR-70-1935 to the California Institute of Technology.

* The author is a Research Fellow at the California Institute of Technology and a member of the technical staff of Applied Theory, Inc., Westwood, Los Angeles, California.

1 Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. I, 214–16.
2 M. G. Krein, Dokl. Akad. Nauk SSSR **76**, 345 (1951).
3 V. A. Ambarzumian, Z. Phyzik **53**, 690 (1929).
4 G. Borg, Acta Math. **78**, 1 (1946).
5 B. M. Levitan and M. G. Gasymov, Russ Math. Surv. **19**, No. 2, 1 (1964).
6 R. E. Eaves, T. G. Twardeck, and S. J. Morin, IEEE Trans. Antennas Propagation **19**, No. 2 (1971).
7 R. E. Langer, Amer. Math. Soc. Bull. **39**, 814 (1933).
8 L. B. Slichter, Physics **4**, 307 (1933).
9 I. M. Gel'fand and B. M. Levitan, Amer. Math. Soc. Transl., Ser. 2, No. 1, 253 (1955).
10 V. A. Marchenko, Tr. Mosk. Natural'n Obs. **1**, 327 (1952); *ibid.* **2**, 3 (1953).
11 L. D. Faddeyev, "The Inverse Problem in the Quantum Theory of Scattering," Research Report No. EM-165, NYU Div. of EM Research (1960).
12 B. Gopinath and M. Sondhi, Proc. IEEE **59**, No. 3, 383 (1971).
13 Irvin Kay, "The Inverse Scattering Problem," Research Report No. EM-74, NYU Div. of EM Research (1955).
14 Irvin Kay, "On the Determination of the Free Electron Distribution of an Ionized Gas," Research Report No. EM-141, NYU Div. of EM Research (1959).
15 Irvin Kay, Commun. Pure Appl. Math. **12**, 371 (1960).
16 G. N. Balanis, "Plasma Inverse Scattering Theory," PhD. Thesis in EE Dept. of California Institution of Tech. (June, 1972).
17 The electron density $N(z)$ is assumed to be a nonnegative, bounded, continuous, and with continuous derivative function of z throughout this paper.
18 It can be easily seen from (14)–(19) that $|C_1|, |C_2| \to 0$ as $k \to \pm \infty$.
19 $H(\xi)$ is the Heaviside step function,
$$H(\xi) = \begin{cases} 0, & \xi < 0 \\ 1, & \xi > 0 \end{cases}.$$
20 It is of interest to note that we could also obtain $k_p^2(z)$ from the characteristic fields by solving for $k_p^2(z)$ in Eqs. (39) and (43); however, such a procedure is rather cumbersome.
21 It is evident that this essential property of the characteristic fields is a consequence of the invariance of the plasma wave equation under a time reversal.
22 S. G. Mikhlin, *Linear Equations of Mathematical Physics* (Holt and Rinehart, New York, 1967), p. 27.

Lattice Wind-Tree Models. I. Absence of Diffusion

D. J. Gates*
Mathematics Department, Rockefeller University, New York, N.Y. 10021
(Received 19 January 1972; Revised Manuscript Received 6 March 1972)

Some lattice versions of the wind-tree model of Ehrenfest are introduced and studied. They include two versions in which overlapping of tree particles is forbidden, a third in which it is allowed, and a fourth in which it is allowed but results in a finite repulsive force. In every case it is found that the mean-square displacement $\Delta(t)$ of the wind particles at time t is bounded independently of t at sufficiently high density of the trees. This is in sharp contrast with the Einstein relation $\Delta(t) = O(t)$ as $t \to \infty$, which might be expected to hold at low densities. Randomization of the initial velocity of a wind particle is also shown to occur in a certain sense, and upper bounds on a recurrence time are obtained at high density. On the other hand, it is shown that thermalization does not occur even at low densities.

1. INTRODUCTION

The Lorentz model and the "wind-tree" model of Ehrenfest have been used recently for the study of the diffusion process and the problem of approach to equilibrium (e.g., Refs. 1, 2, 3). In these models there are two types of particles, called wind particles and tree particles (or trees). The trees are at rest and distributed in equilibrium under the action of their mutual forces and are unaffected by the motion of the wind particles. The latter have no mutual forces, so that their evolution is determined entirely by their interaction (or collisions) with the trees and the container walls.

We consider a system of square trees with sides of length 1, centered on the points (Z^2) of a simple square lattice with unit spacing. A typical configuration is illustrated in Fig. 1 where the lattice is shown inclined at 45°. We specify that the system be enclosed by a box consisting of a close-packed arrangement of trees; but the over-all shape of the box is otherwise arbitrary.

We consider a wind particle which sits initially at some point $\mathbf{q}$ and has initial velocity $\mathbf{p}$ along one of the four lattice directions at 45° to the lattice lines. The wind particle is deflected through a right-angle whenever it strikes a tree, as shown in Fig. 1. Its position $\mathbf{q}_t(\mathbf{x}, C, \Lambda)$ and velocity $\mathbf{p}_t(\mathbf{x}, C, \Lambda)$ at time t depend only on the initial dynamical state $\mathbf{x} \equiv (\mathbf{q}, \mathbf{p})$, the set C

of lattice sites occupied by the centers of the trees which do not comprise the box, and the set Λ of available lattice sites in the box. A wind particle initially on the edge of a tree with velocity directed towards a tree is taken to have its natural velocity of deflection for $t > 0$. A particle which strikes the corner of a tree along a line through the center of the tree is taken to be reflected back along its path on striking the tree. A particle which grazes the corner of a tree is taken to be undeflected. For a particle initially inside a tree, or on the common edge of two touching trees, we choose $\mathbf{q}_t = \mathbf{q}$ and $\mathbf{p}_t = 0$, which means, in effect, that the wind particle is "stuck" inside the tree, or between the trees.

We assign to different sets C the probability distribution

$$f_\Lambda(C) \equiv \begin{cases} z^{|C|}/Z_\Lambda & \text{for } C \subset \Lambda \\ 0 & \text{otherwise} \end{cases}, \tag{1.1}$$

where $z > 0$ is a constant, $|C|$ is the number of points in C, and

$$Z_\Lambda = \sum_{C \subset \Lambda} z^{|C|} = (1+z)^{|\Lambda|}. \tag{1.2}$$

This function $f_\Lambda(C)$ is the grand-canonical distribution for an equilibrium state of the trees with fugacity z, and Z_Λ is the grand partition function. The probability of finding a tree on a given lattice site is simply $z/(1+z)$. The dependence of functions on z is not shown in the notation.

We define the average velocity

$$\bar{\mathbf{p}}_t(\mathbf{x}, \Lambda) \equiv \sum_{C \subset \Lambda} f_\Lambda(C) \mathbf{p}_t(\mathbf{x}, C, \Lambda) \tag{1.3}$$

and mean-square displacement

$$\Delta(t, \mathbf{x}, \Lambda) \equiv \sum_{C \subset \Lambda} f_\Lambda(C) |\mathbf{q}_t(\mathbf{x}, C, \Lambda) - \mathbf{q}|^2 \tag{1.4}$$

of a wind particle moving in the equilibrium system of trees. The functions have well-defined thermodynamic limits ($|\Lambda| \to \infty$) obtained by taking larger and larger boxes which recede from $\mathbf{q}$ in all directions. These limits are given by

$$\bar{\mathbf{p}}_t(\mathbf{x}) = (1+z)^{-|R(t,\mathbf{x})|} \sum_{C \subset R(t,\mathbf{x})} z^{|C|} \mathbf{p}_t(\mathbf{x}, C) \tag{1.5}$$

and

$$\Delta(t, \mathbf{x}) = (1+z)^{-|R(t,\mathbf{x})|} \sum_{C \subset R(t,\mathbf{x})} z^{|C|} |\mathbf{q}_t(\mathbf{x}, C) - \mathbf{q}|^2, \tag{1.6}$$

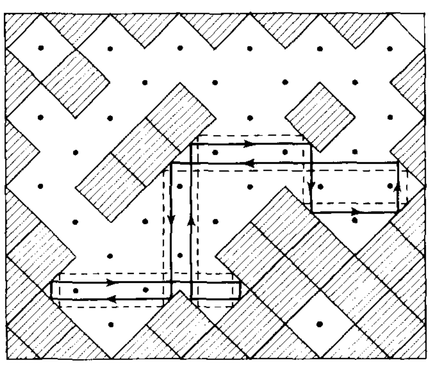

FIG. 1. A possible arrangement of trees in model I. A trajectory and the corresponding central trajectory (broken line) is shown.

where $R(t, \mathbf{x})$ is the set of lattice points upon which the placing of a tree can influence the position of a wind particle up to time t. It contains of order $(|\mathbf{p}|t)^2$ points. As indicated in the notation, $\mathbf{p}_t$ and $\mathbf{q}_t$ are clearly independent of Λ in this limit. To prove (1.5) and (1.6), we simply note that if $\Lambda \supset R$, then from (1.1), the probability of a set C in R is just $z^{|C|}(1+z)^{-|R|}$.

Approach to equilibrium, in a sense, can be studied by considering (as suggested to the author by Professor M. Kac) the long-term behavior of $\bar{\mathbf{p}}_t$. One could say that velocity equilibrium was approached in finite volume if

$$\bar{\mathbf{p}}_t(\mathbf{x}, \Lambda) \to 0 \quad \text{as } t \to \infty \tag{1.7}$$

and similarly in infinite volume. These would imply that the initial velocity is completely randomized when $t \to \infty$ as a result of collisions, and seem intuitively obvious. We find that, on the contrary, they do not hold if the density of the trees is sufficiently high (Sec. 6). The following weakened version of (1.7), in which the limit is replaced by a Cesaro limit, is however obvious

$$\frac{1}{T} \int_0^T dt\, \bar{\mathbf{p}}_t(\mathbf{x}, \Lambda) \to 0 \quad \text{as } T \to \infty, \tag{1.8}$$

because the left side is bounded in magnitude by

$$\begin{aligned} (1/T) &\sum_{C \subset \Lambda} f_\Lambda(C) \left| \int_0^T dt\, \mathbf{p}_t(\mathbf{x}, C, \Lambda) \right| \\ &= (1/T) \sum_{C \subset \Lambda} f_\Lambda(C) |\mathbf{q}_T(\mathbf{x}, C, \Lambda) - \mathbf{q}| \\ &\le d(\Lambda)/T, \end{aligned} \tag{1.9}$$

where $d(\Lambda)$ is the maximum diameter of Λ. We show (Corollary 1) that the analogous result also holds in infinite volume if the density of the trees is sufficiently high.

Diffusion of wind particles can be studied by considering the long-term behavior of the mean-square displacement $\Delta(t, \mathbf{x})$. The work of Hauge and Cohen[1] and Wood and Lado[2] suggests that the particles obey the normal Einstein diffusion law

$$\Delta(t, \mathbf{x}) \approx 2Dt \quad \text{for large } t \tag{1.10}$$

at low enough densities (small z), where D is a number, depending on z, known as the diffusion coefficient. In fact, for the continuum wind-tree model, a result like (1.10) has been shown by Gallavotti[3] to hold asymptotically in a special limit, known as the Boltzmann limit. The effect of this limit is to convert the evolution of the system into a Markov process, of which (1.10) is a well-known consequence.[4]

However, the evolution of the given system (without the Boltzmann limit) is not Markovian, as is well known,[1] and there is no *a priori* reason to believe that (1.10) holds even at low densities. In fact, for the continuum model in which trees are allowed to overlap, both Hauge and Cohen[1] and Wood and Lado[2] predict that $\Delta(t)$ tends to infinity more slowly than t.

It seems possible that similar "abnormal diffusion" will occur also in the nonoverlapping case at sufficiently high densities. Only moderate densities have so far been considered, where (1.10) seems to be confirmed. Abnormal diffusion might also be expected to occur in the present lattice model. We show that this is indeed so. In fact, we prove (Corollary 1) that at high enough densities there is a complete absence of diffusion, in the sense that $\Delta(t, \mathbf{x})$ is bounded by a constant depending only on z.

It is clear that the diffusion process in these models is not like the more common diffusion processes in which the scattering particles also move. It is therefore all the more surprising and interesting that the Einstein law (1.10) seems to hold for them under some circumstances.

Besides the above model (I), we consider in Sec. 3 a model (II) in which trees are allowed to overlap. Again we prove absence of diffusion; but now the result holds at much lower densities (Theorem 2). This is consistent with the abnormality found in the continuum model for overlapping trees.[1,2] In Sec. 4 we consider a model (III) in which the trees form a so-called "hard-squares lattice gas." The statistics of this model are more complex; but we are able to prove essentially the same results (Theorem 3) using some techniques of Dobrushin.[5] In Sec. 5, we consider an intermediate model (IV) in which overlapping is allowed, but results in a finite repulsive force. Absence of diffusion is again proved (Theorem 4). Recurrence times are considered in Sec. 6, and the problem of thermalization is studied in Sec. 7. The connection with percolation theory is discussed in Sec. 8.

We point out that the fugacity z is a measure of the density ρ of the trees. These quantities are related by the formula

$$\rho = \lim_{\Lambda \to \infty} \frac{1}{|\Lambda|} z \frac{\partial}{\partial z} \log Z_\Lambda, \tag{1.11}$$

where Z_Λ is given by (1.2) for models I and II, by (4.2) for model III, and by (5.2) for model IV. It follows[6] that ρ is an increasing function of z. For example, in models I and II one finds that $\rho = z(1+z)^{-1}$; but for models III and IV, ρ is not completely known.

2. MODEL I—NO OVERLAPPING

Our main results can be stated in the following form.

Theorem 1: For model I defined above with $z > 15$, all the moments

$$m_\alpha(t, \mathbf{x}, \Lambda) \equiv \sum_{C \subset \Lambda} f_\Lambda(C) |\mathbf{q}_t(\mathbf{x}, C, \Lambda) - \mathbf{q}|^\alpha \tag{2.1}$$

for any $\alpha > 0$, are bounded uniformly in t, $\mathbf{x}$, and Λ.

The physically interesting consequences are given by

Corollary 1: For model I, the following hold for $z > 15$:

$$\frac{1}{T}\int_0^T dt\, \bar{\mathbf{p}}_t(\mathbf{x}) \to 0 \quad \text{as } T \to \infty \tag{2.2}$$

uniformly in $\mathbf{x}$, and there is a constant A, depending only on z, such that

$$\Delta(t, \mathbf{x}) < A \tag{2.3}$$

for all t and $\mathbf{x}$.

The condition $z > 15$ corresponds to a density $\rho > 15/16$ while the close-packing density is $\rho = 1$. Since these results hold uniformly in $\mathbf{x}$, we can replace $\mathbf{p}_t$ and Δ by their averages with respect to $\mathbf{x}$ over any probability distribution.

The statement (2.3) follows immediately from Theorem 1 and the statement (2.2) follows from the relations (1.9) and Theorem 1.

To prove Theorem 1, we consider *central trajectories*, defined as those trajectories which impinge on trees at the midpoints of their edges, as shown in Fig. 1. By looking at examples of trajectories, one concludes the following.

Lemma: For any trajectory there corresponds a central trajectory which remains always at a distance $< 1/2\sqrt{2}$ from the given trajectory (with the exception of trajectories which lie along the lattice lines).

Roughly speaking, one can say that any trajectory is parallel to a central trajectory. If $\mathbf{x}'$ is the initial dynamical state of a wind particle, on the central trajectory, which has the same velocity $\mathbf{p}$ as the given particle, and is initially at least distance from the given particle, then

$$|\mathbf{q}_t(\mathbf{x}, C, \Lambda) - \mathbf{q}_t(\mathbf{x}', C, \Lambda)| < (1/2\sqrt{2}) \text{ for all } t, C, \text{and } \Lambda.$$

It is therefore sufficient to prove Theorem 1 for the central trajectories.

To analyze the central trajectories, we replace the trees by squares of side $1/\sqrt{2}$, called *subtrees*, as shown in Fig. 2. The extensions of the sides of all possible such squares form another square lattice. This is a horizontal lattice with spacing $1/\sqrt{2}$. The subtrees occupy the spaces between the new lattice lines, and the central trajectories lie always on the lattice lines. Only one sublattice of the squares can be occupied, so that the close-packing arrangement of trees becomes a checkerboard pattern of subtrees. A trajectory never lies along the edge of a tree. The collision law for a particle moving on these lattice lines can easily be seen from Fig. 2.

Because of our choice of boundary conditions, every trajectory J is a closed (possibly self-intersecting) polygon around which a particle circulates for all time. If (J) is the family of sets C which permit the

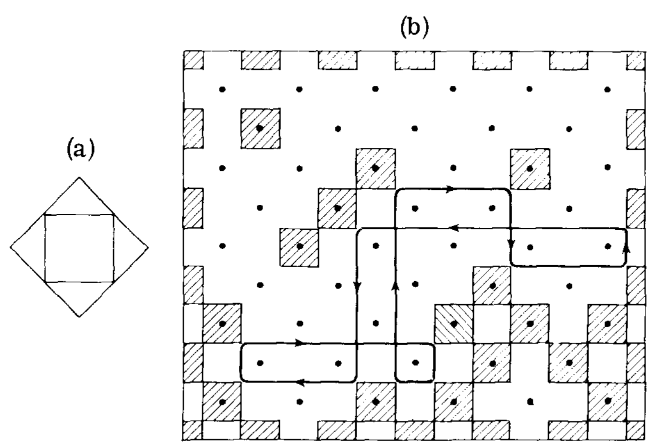

FIG. 2. (a) Construction of the subtree of a tree, and (b) the subtrees and the central trajectory corresponding to Fig. 1.

trajectory J, then the probability $P_\Lambda(J)$ of J is simply

$$P_\Lambda(J) = \sum_{C \in (J)} f_\Lambda(C). \tag{2.4}$$

Let $[\mathbf{x}]$ be the set of all trajectories J that are consistent with the initial state $\mathbf{x}$. Let $\mathbf{q}_t'(\mathbf{x}, J)$ be defined for $J \in [\mathbf{x}]$ as the position at time t of a particle with initial state $\mathbf{x}$ moving on J. Since each set C yields a unique $J \in [\mathbf{x}]$, it follows from (2.1) that

$$m_\alpha(t, \mathbf{x}, \Lambda) = \sum_{J \subset [\mathbf{x}]} \sum_{C \subset (J)} f_\Lambda(C) |\mathbf{q}_t'(\mathbf{x}, J) - \mathbf{q}|^\alpha$$
$$= \sum_{J \subset [\mathbf{x}]} P_\Lambda(J) |\mathbf{q}_t'(\mathbf{x}, J) - \mathbf{q}|^\alpha. \tag{2.5}$$

To estimate $P_\Lambda(J)$, we note that every lattice bond included in J is bordered on one side (only) by an empty square which contains a point in $\mathbf{Z}^2$ (the lattice on which the original trees were placed). At least $\frac{1}{4}$ of this square can be assigned exclusively to the bond. Hence a trajectory J of length j must be bordered by a set of at least $\frac{1}{4}j$ empty sites of $\mathbf{Z}^2$. The probability distribution of sets D on this set, E say, of empty squares is, according to (1.1),

$$f_E(D) = z^{|D|}(1 + z)^{-|E|}.$$

Since $|E| \geq \frac{1}{4}j$, it follows that

$$P_\Lambda(J) \leq f_E(\emptyset) \leq (1 + z)^{-j/4}. \tag{2.6}$$

The number of trajectories of length j in $[\mathbf{x}]$ is less than 2^{j-1}, because a trajectory has at most two possible directions in which to proceed at each vertex of the new lattice (Fig. 2). A particle on J cannot be displaced from $\mathbf{q}$ by a distance greater than $\frac{1}{2}j$, regardless of time or initial velocity. It follows that

$$m_\alpha(t, \mathbf{x}, \Lambda) \leq 2^{-\alpha-1} \sum_j j^\alpha [2(1 + z)^{-1/4}]^j. \tag{2.7}$$

If we let the sum extend to $j = \infty$, this bound holds for all t, $\mathbf{x}$, and Λ. The sum is finite only if $2(1 + z)^{-1/4} < 1$, which proves Theorem 1.

3. MODEL II—OVERLAPPING

A lattice analog of the overlapping trees model considered by Hauge and Cohen[1] is obtained by placing trees of side $\sqrt{2}$ on $\mathbf{Z}^2$. Now we put $\mathbf{Z}^2$ in the horizontal-vertical orientation. (Alternatively we can image that the model I is modified by adding the complementary lattice and allowing all sites to be occupied.) A possible configuration is illustrated in Fig. 3; but no

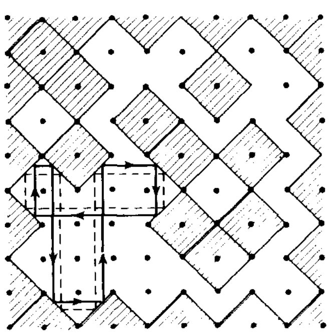

FIG. 3. A possible arrangement of trees in model III showing a trajectory and the corresponding central trajectory (broken line).

overlapping is shown because the diagram is intended for the description of model III. We take the boundary to be formed by a closed packed nonoverlapping array of trees. Equations (1.1)–(1.6) are applicable to this model because the trees again constitute an ideal lattice gas. The main result is

Theorem 2: The moments $m_\alpha(t, \mathbf{x}, \Lambda)$ are bounded uniformly in t, $\mathbf{x}$, and Λ for model II if $z > 8$.

The analog of Corollary 1 follows. The result is stronger than Theorem 1, which is understandable because overlapping implies that many trajectories are returned parallel to themselves. In other words, there is a greater tendency for wind particles to be trapped. This shows the same trend as the results of Refs. 1 and 2. The condition $z > 8$ corresponds to a density $\rho > \frac{8}{9}$, while the close-packing density is $\rho = 1$ and the space-filling density is $\rho = \frac{1}{2}$.

The proof is almost the same as for Theorem 1. The first difference is that a lattice bond included in a central trajectory J must be bordered on *both* sides by an empty square which contains a point of $\mathbf{Z}^2$. Hence a trajectory of length j must be bordered by a set of at least $\frac{1}{2}j$ empty sites of $\mathbf{Z}^2$. The second difference is that central trajectories terminate when they meet a pair of overlapping trees. A wind particle on such a trajectory simply oscillates back and forth, and cannot be displaced from $\mathbf{q}$ by a distance greater than j. The third difference is that a particle can be deflected to the right or left by a tree, so there are less than 3^{j-1} trajectories of length j. Inequality (2.9) is therefore replaced by

$$m_\alpha(t, \mathbf{x}, \Lambda) \leq \tfrac{1}{3} \sum_j j^\alpha [3(1 + z)^{-1/2}]^j \tag{3.1}$$

which proves Theorem 2.

4. MODEL III—HARD-SQUARE LATTICE GAS

A more interesting model is obtained from model II by imposing the condition that no two trees may overlap, so that two neighboring lattice sites cannot be simultaneously occupied. A typical configuration is shown in Fig. 3. This system of trees is known as a *hard-square lattice gas*, and has been extensively studied.[5,7,8] It is known to have a crystal-fluid phase transition. Following Dobrushin,[5] we specify that the system be enclosed by a box formed by a *rectangular* array of close-packed trees which extends to infinity outside the box. The size of the box can be changed only by removing from or adding to the trees in this close-packed configuration. This means that the boundary trees are confined to one of the two sublattices and is essential for our proofs to work. It is also essential that sides of the rectangle of lattice sites occupied by the boundary trees contain an *even number* of sites.

We assign to different sets C of occupied lattice sites the probability distribution

$$f_\Lambda(C) = \begin{cases} z^{|C|}/Z_\Lambda & \text{for } C \in (\Lambda) \\ 0 & \text{otherwise} \end{cases}, \tag{4.1}$$

where (Λ) is the family of allowed sets in Λ taking account of the hard-square repulsion, and

$$Z_\Lambda \equiv \sum_{C \in (\Lambda)} z^{|C|}. \tag{4.2}$$

This partition function Z_Λ is not explicitly calculable. Again $f_\Lambda(C)$ is the grand-canonical distribution for an equilibrium state of the trees with fugacity z. The probability of finding a tree on a given lattice site is not explicitly known, but is certainly not the same for all lattice sites. (This statement is true even when $\Lambda \to \infty$ if z is large, as shown by Dobrushin.[5]) The average density ρ is given in principle by (1.11) and has been studied numerically.[7,8]

The average velocity $\bar{\mathbf{p}}_t(\mathbf{x}, \Lambda)$ and mean-square displacement $\Delta(t, \mathbf{x}, \Lambda)$ are defined again by (1.3) and (1.4); but we are unable to prove that their infinite volume limits exist (except for small z when the theory of fugacity expansions may be used[3,4]). Instead we note that $|\mathbf{p}_t| = |\mathbf{p}|$ and $|\Delta(t)| \leq |\mathbf{p}t|^2$, so that there exists a vector $\bar{\mathbf{p}}_t(\mathbf{x})$ with components

$$\bar{p}_t^i(\mathbf{x}) \equiv \lim_{\Lambda \to \infty} \sup \bar{p}_t^i(\mathbf{x}, \Lambda), \tag{4.3}$$

where $i = 1, 2$ and $\bar{p}_t^i(\mathbf{x}, \Lambda)$ are the components of $\bar{\mathbf{p}}_t(\mathbf{x}, \Lambda)$, and the limit

$$\Delta(t, \mathbf{x}) \equiv \lim_{\Lambda \to \infty} \sup \Delta(t, \mathbf{x}, \Lambda) \tag{4.4}$$

exists. We obtain

Theorem 3: The moments $m_\alpha(t, \mathbf{x}, \Lambda)$ are bounded uniformly in $t, \mathbf{x}$, and Λ for model III if $z > 16$.

This leads to the analog of Corollary 1. No significance should be attached to the fact that this condition is very close to the condition $z > 15$ specified in Theorem 1. The method of proof for model III is rather different, and both estimates are probably very bad. From the expansions of Gaunt and Fisher[7] we note that $z = 16$ corresponds approximately to $\rho = \frac{15}{16}\rho_0$ where $\rho_0 = \frac{1}{2}$ is the close packing density.

We prove Theorem 3 by again considering central trajectories and subtrees as shown in Fig. 4. Dobrushin[5] constructs "boundaries" for any set C of subtrees, by drawing lines along the common edges of every pair of adjacent unoccupied squares as shown in Fig. 4. These boundaries consist in general of many disconnected parts. We denote by $B(\mathbf{x}, C)$ the part (if any) of the boundaries which is connected by some path to the lattice bond which contains the point $\mathbf{q}$ and has the direction of $\mathbf{p}$. It is clear that the wind particle follows some closed trajectory $(J(\mathbf{x}, C)$ contained in $B(\mathbf{x}, C)$. In fact, J is uniquely determined by B: The rule for finding J in B is to go straight ahead whenever possible; when it is not possible, there is no choice because the boundaries have no T-junctions. Any such boundary B can be completely covered by a single, non-self-intersecting (but possibly self-touching) polygon, called a *contour*, as exemplified in Fig. 5. In general, there are many ways of choosing such a contour on B. Any such contour on B is isolated from other contours not on B because, by definition, B is not connected to any other boundary.

The probability $\Pi_\Lambda(G)$ of finding some specific isolated contour G among all configurations C of the trees is given by

$$\Pi_\Lambda(G) = \sum_{C \in (G)} f_\Lambda(C), \tag{4.5}$$

where (G) is the family of sets C in (Λ) that contain the contour G. Let $\{\mathbf{x}\}$ be the set of contours G that contain the lattice bond through $\mathbf{q}$ in the direction of $\mathbf{p}$. Let $\tilde{\mathbf{q}}_t(\mathbf{x}, G)$ be the position of the wind particle at time t given a contour $G \in \{\mathbf{x}\}$. This position is uniquely determined as pointed out above. From (2.1) and (4.5) we deduce that

$$m_\alpha(t, \mathbf{x}, \Lambda) \leq \sum_{G \in \{\mathbf{x}\}} \Pi_\Lambda(G) |\tilde{\mathbf{q}}_t(\mathbf{x}, G) - \mathbf{q}|^\alpha. \tag{4.6}$$

The inequality results from the fact that there are many ways of drawing a contour G on a given boundary B.

If $\Pi'_\Lambda(G)$ is the probability of G without the restriction that G be isolated, it is clear that $\Pi_\Lambda(G) \leq \Pi'_\Lambda(G)$. Under the boundary conditions stated above, Dobrushin[5] obtained the important result

$$\Pi'_\Lambda(G) \leq z^{-g/4}, \tag{4.7}$$

where g is the length of the perimeter of G. A particle moving on a trajectory in G (see Figs. 4 and 5) is never displaced from $\mathbf{q}$ by a distance greater than $\frac{1}{2}g$ regardless of the shape of G, the initial speed, or the time. Furthermore, there are at most 2^{g-1} contours of length g in $\{\mathbf{x}\}$, because the contours never have two successive segments with the same direction (Fig. 5).

Substituting these estimates in (4.6) yields

$$m_\alpha(t, \mathbf{x}, \Lambda) \leq 2^{-\alpha-1} \sum_g g^\alpha (2z^{-1/4})^g. \tag{4.8}$$

The sum is finite for $2z^{-1/4} < 1$, which proves the theorem.

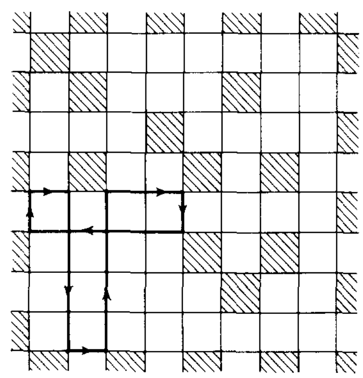

FIG. 4. The subtrees and the central trajectory corresponding to Fig. 3. The Dobrushin boundaries are the lines that separate unoccupied squares.

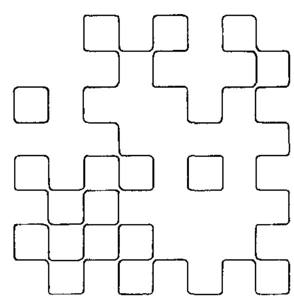

FIG. 5. One possible way of drawing contours on each of the three components of the boundary in Fig. 4. The large contour is G. Note that the path followed by G is not the same as the path followed by the wind particle.

5. MODEL IV—OVERLAPPING AND REPULSION

A model which is intermediate between models II and III contains trees for which overlapping is allowed, but results in a finite repulsive force. The trees form a lattice gas with nearest-neighbor repulsion, of strength $a > 0$ say, which was also considered by Dobrushin. Here the trees are in effect soft with respect to their mutual forces, but hard as far as the wind particles are concerned. The "walls" of the box Λ are formed by a close-packed, nonoverlapping array of trees which have the same repulsive interactions (see Ref. 5 for details).

The sets C of trees have the equilibrium distribution

$$f_\Lambda(C) \equiv \begin{cases} z^{|C|}e^{-\beta U(C)}/Z_\Lambda & \text{for } C \subset \Lambda \\ 0 & \text{otherwise} \end{cases}, \tag{5. 1}$$

where $U(C)$ is the total energy (i.e., strength) of the interactions and is given by a times the number of neighboring pairs of trees in the set including C and the trees in the walls. Also $\beta > 0$ is the reciprocal temperature of the trees and

$$Z_\Lambda \equiv \sum_{C\subset\Lambda} z^{|C|}e^{-\beta U(C)}. \tag{5. 2}$$

Infinite volume limits of $\bar{\mathbf{p}}_t$ and Δ are defined as in model III.

The result we obtain is

Theorem 4: The moments $m_\alpha(t, \mathbf{x}, \Lambda)$ are bounded uniformly in t, $\mathbf{x}$, and Λ for model IV provided that

$$\beta(a - |\tfrac{1}{2}\mu - a|) > 2 \log 3, \tag{5. 3}$$

where $\mu = \beta^{-1} \log z$ is the chemical potential.

For $\mu > 0$, the left side of (5. 3) does not exceed $\frac{1}{2}\beta\mu$, so that this is a stronger result than Theorem 3. However, it is very weak compared to Theorem 2 even for small a. This is due simply to the weakness of our estimates as explained at the end of this section.

The method of proof is identical to that used for Theorem 3. Now, however, the boundaries on the transformed lattice are obtained by drawing lines along the common edges of adjacent empty squares (as before) *and* adjacent occupied squares. The probability $\Pi_\Lambda(G)$ of an isolated contour G covering such a boundary satisfies

$$\Pi_\Lambda(G) \le \exp[\tfrac{1}{2}\beta(|\tfrac{1}{2}\mu - a| - a)g] \tag{5. 4}$$

as shown by Dobrushin.[5] Following the argument leading to (4. 8) yields the statement of Theorem 4.

This result is weak because we have ignored the fact that trajectories end when they meet a pair of overlapping trees. (In fact, Theorem 4 also holds for a model in which wind particles are allowed to pass straight through a pair of overlapping trees.) We might obtain a better result if we could adapt the method of Sec. 3; but it is difficult to estimate the probability of the empty set E in model IV.

6. RECURRENCE TIMES AND LONG TERM LIMITS

We can define a recurrence time $\tau(\Lambda)$ as the average, with respect to tree configurations, of the time taken for a wind particle to return to its starting point with its initial direction. We confine our attention to central trajectories in this section. We also choose $|\mathbf{p}| = 1$ and omit $\mathbf{x}$ dependence from the notation. A particle on a closed trajectory of length $|J|$ returns to its initial state in time $|J|$. Since models I and III contain only closed trajectories it follows that in these models

$$\tau(\Lambda) = \sum_{J\in[\mathbf{x}]} P_\Lambda(J)|J|. \tag{6. 1}$$

Models II and IV contain both closed central trajectories and terminating central trajectories. If we define $|J|$ to be twice the length of J if J is terminating, then (6. 1) again gives the recurrence time.

It is natural to define an infinite volume recurrence time by

$$\tau \equiv \sum_{J\in[\mathbf{x}]} P(J)|J|, \tag{6. 2}$$

the sum being over all closed arbitrarily long J, and where $P(J)$, the probability of J, is equal to the probability of the set of occupied sites in the "sausage" formed by the squares which border J. Two questions immediately arise. Is the infinite series convergent and is it equal to the infinite volume limit of $\tau(\mathbf{x}, \Lambda)$? The answers to both questions are yes for models I and II under our previous high density conditions with a suitable ordering of terms, as we now show; but in general the question remains open.

First we define

$$\tau_-(\Lambda) \equiv \sum_{J\in[\mathbf{x},\Lambda]_-} P_\Lambda(J)|J|, \tag{6. 3}$$

where $[\mathbf{x}, \Lambda]_-$ is the set of trajectories in $[\mathbf{x}]$ which *do not* touch the boundary trees of Λ. For model I we can replace $P_\Lambda(J)$ by $P(J)$ in (6. 3) [but not in (6. 1)]. Since enlarging Λ simply results in adding terms to (6. 3) it follows that

$$\Lambda' \supset \Lambda \quad \text{implies} \quad \tau_-(\Lambda') \ge \tau_-(\Lambda). \tag{6. 4}$$

But the arguments of Sec. 2 imply

$$\tau_-(\Lambda) \le \tfrac{1}{2} \sum_j j[2(1+z)^{-1/4}]^j, \tag{6. 5}$$

so that $\tau_-(\Lambda)$ has a limit as $\Lambda \to \infty$ by means of an ascending sequence of Λ's, provided $z > 15$. But the τ_- are just the partial sums of the series (6. 2), so that their limit is just τ. If we define $[\mathbf{x}, \Lambda]$ as the set of trajectories in $[\mathbf{x}]$ which *do* touch the boundary trees of Λ, we obtain

$$0 \le \tau(\Lambda) - \tau_-(\Lambda) = \sum_{J\in[\mathbf{x},\Lambda]} P_\Lambda(J)|J| \le \tfrac{1}{2} \sum_{j\ge 2d(\mathbf{q},\Lambda)} j[2(1+z)^{-1/4}]^j, \tag{6. 6}$$

where $d(\mathbf{q}, \Lambda)$ is the least distance from $\mathbf{q}$ to the boundary. If $d(\mathbf{q}, \Lambda) \to \infty$ as $\Lambda \to \infty$ and $z > 15$, it follows that $\tau(\Lambda)$ tends also to the same limit τ. The same argument works for model II.

Furthermore, τ has the upper bound given by the right side of (6. 5). Similar upper bounds on $\tau(\Lambda)$ apply to the other models. The same arguments can be applied to the moments

$$M_\alpha(\Lambda) \equiv \sum_{J\in[\mathbf{x}]} P_\Lambda(J)|J|^\alpha \tag{6.7}$$

with $\alpha > 0$. These results immediately raise the question: are the M_α finite when z is small and the density of the trees consequently low?

The second question we consider in this section is the long-term ($t \to \infty$) behavior of the infinite volume moments $m_\alpha(t)$ and the average velocity $\bar{p}_t$ in models I and II. We show that the Cesaro limits

$$\overline{m}_\alpha \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} m_\alpha(t) \tag{6.8}$$

and

$$\tilde{m}_\alpha \equiv \lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, m_\alpha(t) \tag{6.9}$$

exist under the previous high density conditions, and are given by

$$\overline{m}_\alpha = \sum_{J\in[\mathbf{x}]} P(J) \frac{1}{|J|} \sum_{t=1}^{|J|} |\mathbf{q}'_t(J)|^\alpha \tag{6.10}$$

and

$$\tilde{m}_\alpha = \sum_{J\in[\mathbf{x}]} P(J) \frac{1}{|J|} \int_0^{|J|} dt |\mathbf{q}'_t(J)|^\alpha \tag{6.11}$$

with $q'_t(J)$ defined as in Sec. 2.

To prove this we note from the argument used for the M_α that

$$m_\alpha(t) = \sum_{J\in[\mathbf{x}]} P(J)|\mathbf{q}'_t(J)|^\alpha \tag{6.12}$$

under the high density conditions. This shows, by the way, that $m_\alpha(t)$ does not itself tend to a limit as $t \to \infty$, because it is a sum of periodic terms with increasing periods. We deduce that

$$\frac{1}{T} \sum_{t=1}^{T} m_\alpha(t) = \sum_{J\in[\mathbf{x}]} P(J) \frac{1}{T} \sum_{t=1}^{T} |\mathbf{q}'_t(J)|^\alpha. \tag{6.13}$$

But we have $|\mathbf{q}'_t(J)| < |J|/2$, so that

$$\frac{1}{T} \sum_{t=1}^{T} |\mathbf{q}'_t(J)|^\alpha < (|J|/2)^\alpha. \tag{6.14}$$

But we know that the series

$$\sum_{J\in[\mathbf{x}]} P(J)(|J|/2)^\alpha = 2^{-\alpha} M_\alpha \tag{6.15}$$

converges. It follows from the Weierstrass M test that (6.13) converges uniformly in T. We can therefore take the limit $T \to \infty$ term by term in (6.13), which proves (6.10). A similar argument proves (6.11).

One can use the M test to show that the average velocity can be written in the form

$$\bar{p}_t = \sum_{J\in[\mathbf{x}]} P(J)\mathbf{p}'_t(J), \tag{6.16}$$

where $\mathbf{p}'_t(J)$ is the velocity at time t for a particle on J. But the terms in this sum are periodic with increasing periodicity, so that the limit does not exist as $t \to \infty$, contrary to intuition. In fact $\bar{\mathbf{p}}_t$ is clearly an almost periodic function. This behavior is probably peculiar to high densities.

7. ABSENCE OF THERMALIZATION

We consider a system of wind particles of unit mass in equilibrium with fugacity $2\sqrt{2\pi}$ and temperature 1, under the action of an external field $h(\mathbf{q}) \geq 0$ acting only on a finite region. The system is in a box Λ (defined previously for each model) and contains a set C of trees. Following Gallavotti,[3] the one-particle distribution function at time $t = 0$ is given, for $\mathbf{q}$ in the box Λ and $\mathbf{p}$ in one of the four allowed directions, by

$$\rho_C(0, \mathbf{x}, \Lambda) = \exp[-\tfrac{1}{2}\mathbf{p}^2 - h(\mathbf{q})]\chi_C(\mathbf{q}), \tag{7.1}$$

where $\chi_C(\mathbf{q})$ is 0 if $\mathbf{q}$ is inside a tree in the set C, and is 1 otherwise. At time $t = 0$ the field is switched off and the system allowed to evolve. The distribution function at time t is therefore

$$\begin{aligned}\rho_C(t, \mathbf{x}, \Lambda) &= \rho_C[0, \mathbf{x}_{-t}(\mathbf{x}, C, \Lambda), \Lambda] \\ &= \exp\{-\tfrac{1}{2}p^2 - h[\mathbf{q}_{-t}(\mathbf{x}, C, \Lambda)\}\chi_C(\mathbf{q})\end{aligned} \tag{7.2}$$

where $\mathbf{x}_t \equiv (\mathbf{q}_t, \mathbf{p}_t)$ while $\mathbf{q}_t$ and $\mathbf{p}_t$ were defined previously and refer to the trajectory in zero field. The second equality in (7.2) holds because the kinetic energy, the set of allowed $\mathbf{q}$, and the set of disallowed $\mathbf{q}$ are all conserved by the motion. The latter results from our condition that a particle initially in a tree remains "stuck" in that tree for all time.

The average of ρ_C is defined by

$$\begin{aligned}\rho(t, \mathbf{x}, \Lambda) = \exp(-\tfrac{1}{2}\mathbf{p}^2) \sum_{C\subset\Lambda} f_\Lambda(C)\chi_C(\mathbf{q}) \\ \times \exp\{-h[\mathbf{q}_{-t}(\mathbf{x}, C, \Lambda]\}.\end{aligned} \tag{7.3}$$

We say that the system thermalizes if $\rho(t, \mathbf{x}, \Lambda)$ tends, as $\Lambda \to \infty$ and $t \to \infty$, to the equilibrium distribution function in zero field averaged over tree configurations, namely

$$\begin{aligned}\rho^0(\mathbf{x}, \Lambda) &= \exp(-\tfrac{1}{2}\mathbf{p}^2) \sum_{C\subset\Lambda} f_\Lambda(C)\chi_C(\mathbf{q}) \\ &= \exp(-\tfrac{1}{2}\mathbf{p}^2) \sum_{C\subset\Lambda\setminus S(\mathbf{q})} f_\Lambda(C),\end{aligned} \tag{7.4}$$

where $S(\mathbf{q})$ is the set of (0, 1, or 2) lattice points upon which the placing of a tree would cover a wind particle at $\mathbf{q}$. The sum in (7.4) is just the probability that $\mathbf{q}$ is not inside a tree.

As a diversion we note from (3.1) that, for models I and II,

$$\rho^0(\mathbf{x}, \Lambda) = \exp(-\tfrac{1}{2}\mathbf{p}^2)(1 + z)^{-|S(\mathbf{q})|} \tag{7.5}$$

for $\mathbf{q} \in \Lambda$, where $|S(\mathbf{q})| = 1$ in model I and 2 in model II, for almost all $\mathbf{q}$.

We now show that thermalization does *not* occur, using essentially the argument of Gallavotti.[3] Let ω be the (square) region which would be covered if a tree were placed at the origin. Suppose that $h(\mathbf{q}) \geq h' > 0$ for $\mathbf{q} \in \omega$. Then the result we obtain is

$$\rho^0(\mathbf{x}, \Lambda) - \rho(t, \mathbf{x}, \Lambda) \geq \exp(-\tfrac{1}{2}\mathbf{p}^2)[1 - \exp(-h')]\sigma_\Lambda \tag{7.6}$$

for $\mathbf{q} \in \omega$, all $\mathbf{p}$, t, and Λ, where σ_Λ is the probability that the origin contains no tree and ω is bordered by trees, as shown in Fig. 6. For example, in model I, σ_Λ is the probability of finding trees only on the four corner sites in a square array of five sites. Using (1.1) therefore gives $\sigma_\Lambda = z^4(1+z)^{-5}$ for all Λ. In model II we find $\sigma_\Lambda = z^4(1+z)^{-9}$.

It is apparant that $\sigma_\Lambda > \epsilon > 0$ for all Λ in models III and IV also, so that (7.6) states that ρ is bounded away from ρ^0 uniformly in t, Λ, and $q \in \omega$. This excludes the possibility of thermalization. It is a more explicit version of the result obtained in Ref. 3 for the continuum wind-tree model with overlapping trees.

To prove (7.6), we use (7.4) and (7.5) to obtain

$$\rho^0(\mathbf{x},\Lambda) - \rho(t,\mathbf{x},\Lambda) \geq \exp(-\tfrac{1}{2}\mathbf{p}^2) \sum_{C \in \mathfrak{F}} f_\Lambda(C)\chi_C(\mathbf{q}) \times (1-\exp\{-h[\mathbf{q}_{-t}(\mathbf{x},C,\Lambda)]\}), \quad (7.7)$$

where $\mathfrak{F}$ is the family of sets C containing the configuration in Fig. 6. If $\mathbf{q} \in \omega$ and $C \in \mathfrak{F}$, it is clear that the particle is trapped in ω, so that $\mathbf{q}_{-t}(\mathbf{x},C,\Lambda) \in \omega$ for all t. Consequently $h[q_{-t}(\mathbf{x},C,\Lambda)] \geq h'$ for all t and $C \in \mathfrak{F}$, which leads to (7.6).

The infinite volume ($\Lambda \to \infty$) limits of the functions $\rho^0(x,\Lambda)$ and $\rho(t,\mathbf{x},\Lambda)$ can be proved to exist for small enough z, in all four models, by the argument of Ref. 3. They also have power series expansions in z which are convergent for small z. In models I and II, one has the stronger results that

$$\rho^0(\mathbf{x}) \equiv \lim_{\Lambda\to\infty} \rho^0(\mathbf{x},\Lambda) = \exp(-\tfrac{1}{2}\mathbf{p}^2)(1+z)^{-|S(\mathbf{q})|}$$

$$\rho(t,\mathbf{x}) \equiv \lim_{\Lambda\to\infty} \rho(t,\mathbf{x},\Lambda) = \exp(-\tfrac{1}{2}\mathbf{p}^2)(1+z)^{-|R(t,\mathbf{x})|} \times \sum_{C\subset R(t,\mathbf{x})} z^{|C|}\chi_C(\mathbf{q}) \exp\{-h[\mathbf{q}_{-t}(\mathbf{x},C)]\}, \quad (7.8)$$

where $R(t,\mathbf{x})$ was defined in Sec. 1. These limits clearly exist for all z and have power series expansions in z ("fugacity expansions") which converge for $|z| < 1$. These are known however[1,3] to be useless for describing the $t \to \infty$ behavior of $\rho(t,\mathbf{x})$ [and $\Delta(t,\mathbf{x},)$] because the coefficients diverge when $t \to \infty$.

Absence of thermalization immediately raises the question—to what limit, if any, does $\rho(t,\mathbf{x})$ tend when $t \to \infty$? The answer is that it does not tend to any limit under the high density conditions. To show this we use previous arguments to show that

$$\rho(t,\mathbf{x}) = \exp(-\tfrac{1}{2}\mathbf{p}^2) \sum_{J\in[\mathbf{x}]} P(J)\delta_J(\mathbf{q}) \exp\{-h[\mathbf{q}'_{-t}(\mathbf{x},J)]\}, \quad (7.9)$$

where $\delta_J(\mathbf{q})$ is 0 if $\mathbf{q}$ lies inside one of the trees in the sausage surrounding J and is 1 otherwise. Thus $\rho(t,\mathbf{x})$ is a sum of periodic terms with increasing periods, and is therefore almost periodic.

On the other hand our previous arguments also show that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \rho(t,\mathbf{x})$$

exists and is equal to

$$\exp(-\tfrac{1}{2}\mathbf{p}^2) \sum_{J\in[\mathbf{x}]} P(J)\delta_J(\mathbf{q}) \frac{1}{|J|} \sum_{t=1}^{|J|} \exp\{-h[\mathbf{q}'_{-t}(\mathbf{x},J)]\}. \quad (7.10)$$

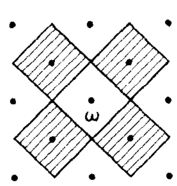

FIG. 6. The configuration of trees which traps a wind particle in ω.

8. DISCUSSION

The models discussed here are to some extent related[9] to models of percolation.[10-12] In the latter models bonds are removed from a lattice and quantities like the probability of an infinite connected piece of lattice are studied. A walk on such a lattice is reminiscent of a trajectory in our models if we take the presence of a subtree (see Fig. 4) to represent the cutting of four bonds. In our models, however, there are strict dynamical laws which influence the results. For example, in model I, the number of trajectories of length j is bounded as 2^{j-1}; but the number of walks of the same length is bounded only as 3^{j-1}.

The results we have obtained give rise to many interesting questions and have many possible extensions and generalizations. One would expect that Theorem 4 could be strengthened by using a method like that used in the proof of Theorem 2, where the effect of overlapping is taken into account. It should be possible, in turn, to strengthen Theorems 1 and 2 by a more refined analysis. Our methods certainly give us no reason to believe that our estimates are close to the best possible ones.

It may be possible to extend our results to the countless variants of the present models, with trees of various sizes on different types of lattice. For example, we could have hexagonal trees on a triangular lattice (see Ref. 5), where the wind particles can move in the six possible lattice directions and are deflected through 60° by a tree. Analogues of the four models considered in the present paper could be studied. Three-dimensional models might also be of interest.

The most important problem raised by our results is to determine the low density behavior of the models. The first thing one might try to prove is that, at least, $\Delta(t) \to \infty$ as $t \to \infty$ for small enough z. This in itself would be interesting, because it would imply the existence of a transition in the diffusion behavior for some value of z. In model III, this transition may coincide with the freezing transition in the trees; but our results do not indicate this: our upper bound 16 on z in Theorem 3 is lower than that required by Dobrushin to ensure that the trees are in their "frozen" state. However, no definite conclusion can be drawn, because the freezing transition has been shown numerically[4] to occur when $z \approx 4$. Furthermore, there is no freezing transition in models I and II.

The absence of thermalization proved in Sec. 7 shows that the effect of trapping is felt even at low densities. It is natural to ask to what extent thermalization, or its absence, is related to diffusion, or its absence. Suppose $F_t(r)$ is the probability that a wind particle travels a distance not exceeding r during time t. From Chebychev's inequality we deduce that

$$\Delta(t) \geq r^2[1 - F_t(r)] \quad \text{for all } r. \quad (8.1)$$

Thermalization means that $F_t(r) \to 0$ as $t \to \infty$, which implies that $\Delta(\infty) > r^2$ for any r. Thermalization therefore implies existence $[\Delta(t) \to \infty]$ of diffusion. Conversely, absence of diffusion $[\Delta(t) < A]$ implies absence of thermalization. However, absence of thermalization (which we have) does not exclude diffusion, and even allows normal diffusion $[\Delta(t)/t \to 2D > 0]$. This can be seen by considering a density function $f_t(r)$, for a wind particle to travel a distance r in time t, given by

$$f_t(r) = \begin{cases} a_t & \text{for } r \le 1 \\ a_t r^{-3-(1/t)} & \text{for } r \ge 1 \end{cases}, \tag{8.2}$$

where $a_t \equiv (1 + 2t)/(1 + 3t)$. As $t \to \infty$, this density function tends to the limit

$$f_\infty(r) = \begin{cases} \frac{2}{3} & \text{for } r \le 1 \\ \frac{2}{3} r^{-3} & \text{for } r \ge 1 \end{cases} \tag{8.3}$$

which implies absence of thermalization. However

$$\frac{\Delta(t)}{t} = \frac{1}{t} \int_0^\infty dr r^2 f_t(r) = \frac{(1+2t)}{3t} \to \tfrac{2}{3} \quad \text{as } t \to \infty, \tag{8.4}$$

which is normal diffusion. [Try also replacing t by t^n in (8.2)!]

It is clear that most of the important questions remain unanswered. An indication of these answers may be obtainable from computer experiments like those of Wood and Lado.

ACKNOWLEDGMENTS

I am greatly indebted to Professor E. G. D. Cohen and Professor M. Kac for introducing me to this field and to them, Professor J. T. Cannon, and Professor J. L. Lebowitz for valuable discussions.

* Supported by the U.S. Air Force under Grant No. AFOSR 72-2187.

1 E. H. Hauge and E. G. D. Cohen, J. Math. Phys. **10**, 397 (1969).
2 W. W. Wood and F. Lado, J. Comput. Phys. **7**, 528 (1971).
3 G. Gallavotti, Phys. Rev. **185**, 308 (1969).
4 *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954).
5 R. L. Dobrushin, Functional Anal. Appl. 2, No. 4, 44 (1968); [English translation 2, 302 (1968)]. An excellent review is given by J. Ginibre, *Colloques Internationaux du CNRS*, No. 181 (Editions du CNRS, Paris, 1970), p. 163.
6 M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964). See also D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), Chap. 2.
7 D. S. Gaunt and M. E. Fisher, J. Chem. Phys. **43**, 2840 (1965).
8 L. K. Runnels and L. L. Combes, J. Chem. Phys. **45**, 2482 (1966).
9 I am grateful to Professor J. L. Lebowitz for this observation.
10 M. E. Fisher and J. W. Essam, J. Math. Phys. **2**, 609 (1961).
11 M. E. Fisher, J. Math. Phys. **2**, 620 (1961).
12 C. M. Fortuin, "On the Random Cluster Model" (to be published).

General Theory of Spherically Symmetric Boundary-Value Problems of the Linear Transport Theory*

Madhoo Kanal
Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01002
(Received 18 June 1971; Revised Manuscript Received 31 January 1972)

A general theory of spherically symmetric boundary-value problems of the one-speed neutron transport theory is presented. The formulation is also applicable to the "gray" problems of radiative transfer. The Green's function for the purely absorbing medium is utilized in obtaining the normal mode expansion of the angular densities for both interior and exterior problems. As the integral equations for unknown coefficients are regular, a general class of reduction operators is introduced to reduce such regular integral equations to singular ones with a Cauchy-type kernel. Such operators then permit one to solve the singular integral equations by the standard techniques due to Muskhelishvili. We discuss several spherically symmetric problems. However, the treatment is kept sufficiently general to deal with problems lacking azimuthal symmetry. In particular the procedure seems to work for regions whose boundary coincides with one of the coordinate surfaces for which the Helmholtz equation is separable.

INTRODUCTION

In an earlier paper Case *et al.*[1] solved several spherically symmetric boundary-value problems of the one-speed neutron transport theory for homogeneous isotropically scattering media. In particular they illustrate the Green's function approach, introduced by Case,[2] by directly obtaining the normal mode expansions of the angular densities for interior and exterior problems. They also show that the appropriate integral equations for expansion coefficients, which are regular (due to the regularity of the normal modes) and hard to solve by conventional methods, may be replaced by soluble auxiliary singular integral equations for the same expansion coefficients. Thus, they introduce certain reduction operators which have essentially the property of transforming such regular integral equations into singular integral equations with a Cauchy-type kernel. In this paper we extend their treatment in two ways. First we remove the restriction that the media are isotropically scattering which entails the introduction of two general classes of reduction operators corresponding to interior and exterior problems. Second, we introduce a slightly different method which utilizes the infinite-medium Green's function G_0 for a purely absorbing medium in formulating the boundary-value problems. Regarding this second aspect, we show that, although the final integral equations one must solve for expansion coefficients are identical to the ones encountered in the formulation utilizing the infinite-medium Green's function G characteristic of the medium, our approach has the advantage that it eliminates the cumbersome process of constructing different Green's functions for media with different scattering properties. In that sense the methodology presented here incorporates the Green's function technique discussed in Ref. 2 and illustrated for spherical geometry in Ref. 1. More specifically, we introduce certain restricted Fourier amplitudes for any arbitrary geometry. Thus, if one expands the scattering kernel in terms of say, spherical harmonics, then those restricted implies that $\Delta(\infty) > r^2$ for any r. Thermalization therefore implies existence $[\Delta(t) \to \infty]$ of diffusion. Conversely, absence of diffusion $[\Delta(t) < A]$ implies absence of thermalization. However, absence of thermalization (which we have) does not exclude diffusion, and even allows normal diffusion $[\Delta(t)/t \to 2D > 0]$. This can be seen by considering a density function $f_t(r)$, for a wind particle to travel a distance r in time t, given by

$$f_t(r) = \begin{cases} a_t & \text{for } r \le 1 \\ a_t r^{-3-(1/t)} & \text{for } r \ge 1 \end{cases}, \tag{8.2}$$

where $a_t \equiv (1 + 2t)/(1 + 3t)$. As $t \to \infty$, this density function tends to the limit

$$f_\infty(r) = \begin{cases} \frac{2}{3} & \text{for } r \le 1 \\ \frac{2}{3} r^{-3} & \text{for } r \ge 1 \end{cases} \tag{8.3}$$

which implies absence of thermalization. However

$$\frac{\Delta(t)}{t} = \frac{1}{t} \int_0^\infty dr r^2 f_t(r) = \frac{(1 + 2t)}{3t} \to \tfrac{2}{3} \quad \text{as } t \to \infty, \tag{8.4}$$

which is normal diffusion. [Try also replacing t by t^n in (8.2)!]

It is clear that most of the important questions remain unanswered. An indication of these answers may be obtainable from computer experiments like those of Wood and Lado.

ACKNOWLEDGMENTS

I am greatly indebted to Professor E. G. D. Cohen and Professor M. Kac for introducing me to this field and to them, Professor J. T. Cannon, and Professor J. L. Lebowitz for valuable discussions.

* Supported by the U.S. Air Force under Grant No. AFOSR 72-2187.

1 E. H. Hauge and E. G. D. Cohen, J. Math. Phys. **10**, 397 (1969).
2 W. W. Wood and F. Lado, J. Comput. Phys. 7, 528 (1971).
3 G. Gallavotti, Phys. Rev. **185**, 308 (1969).
4 *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954).
5 R. L. Dobrushin, Functional Anal. Appl. 2, No. 4, 44 (1968); [English translation 2, 302 (1968)]. An excellent review is given by J. Ginibre, *Colloques Internationaux du CNRS*, No. 181 (Editions du CNRS, Paris, 1970), p. 163.
6 M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964). See also D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), Chap. 2.
7 D. S. Gaunt and M. E. Fisher, J. Chem. Phys. **43**, 2840 (1965).
8 L. K. Runnels and L. L. Combes, J. Chem. Phys. **45**, 2482 (1966).
9 I am grateful to Professor J. L. Lebowitz for this observation.
10 M. E. Fisher and J. W. Essam, J. Math. Phys. **2**, 609 (1961).
11 M. E. Fisher, J. Math. Phys. **2**, 620 (1961).
12 C. M. Fortuin, "On the Random Cluster Model" (to be published).

General Theory of Spherically Symmetric Boundary-Value Problems of the Linear Transport Theory*

Madhoo Kanal

Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01002

(Received 18 June 1971; Revised Manuscript Received 31 January 1972)

A general theory of spherically symmetric boundary-value problems of the one-speed neutron transport theory is presented. The formulation is also applicable to the "gray" problems of radiative transfer. The Green's function for the purely absorbing medium is utilized in obtaining the normal mode expansion of the angular densities for both interior and exterior problems. As the integral equations for unknown coefficients are regular, a general class of reduction operators is introduced to reduce such regular integral equations to singular ones with a Cauchy-type kernel. Such operators then permit one to solve the singular integral equations by the standard techniques due to Muskhelishvili. We discuss several spherically symmetric problems. However, the treatment is kept sufficiently general to deal with problems lacking azimuthal symmetry. In particular the procedure seems to work for regions whose boundary coincides with one of the coordinate surfaces for which the Helmholtz equation is separable.

INTRODUCTION

In an earlier paper Case *et al.*[1] solved several spherically symmetric boundary-value problems of the one-speed neutron transport theory for homogeneous isotropically scattering media. In particular they illustrate the Green's function approach, introduced by Case,[2] by directly obtaining the normal mode expansions of the angular densities for interior and exterior problems. They also show that the appropriate integral equations for expansion coefficients, which are regular (due to the regularity of the normal modes) and hard to solve by conventional methods, may be replaced by soluble auxiliary singular integral equations for the same expansion coefficients. Thus, they introduce certain reduction operators which have essentially the property of transforming such regular integral equations into singular integral equations with a Cauchy-type kernel. In this paper we extend their treatment in two ways. First we remove the restriction that the media are isotropically scattering which entails the introduction of two general classes of reduction operators corresponding to interior and exterior problems. Second, we introduce a slightly different method which utilizes the infinite-medium Green's function G_0 for a purely absorbing medium in formulating the boundary-value problems. Regarding this second aspect, we show that, although the final integral equations one must solve for expansion coefficients are identical to the ones encountered in the formulation utilizing the infinite-medium Green's function G characteristic of the medium, our approach has the advantage that it eliminates the cumbersome process of constructing different Green's functions for media with different scattering properties. In that sense the methodology presented here incorporates the Green's function technique discussed in Ref. 2 and illustrated for spherical geometry in Ref. 1. More specifically, we introduce certain restricted Fourier amplitudes for any arbitrary geometry. Thus, if one expands the scattering kernel in terms of say, spherical harmonics, then those restricted amplitudes are defined as the volume restricted Fourier transforms of the harmonic moments of the angular density, where the volume is the region in which the angular density is to be determined. It turns out that such amplitudes are precisely the phase space integrals involving Fourier coefficients of G and restricted in the configuration space to the bounding surface of the volume under consideration. Before we elaborate on these two aspects of the restricted amplitudes and how they may be used in obtaining the spectral representations of the angular densities, we wish to point out that in the sequel the treatment of boundary-value problems presented here differs from a similar type given by Case and Zweifel[3] as follows. They utilize G_0 (the Green's function for a purely absorbing medium) to obtain spectral representations of the harmonic moments of the angular density. Their procedure for anisotropic scattering kernels results in a set of coupled integral equations which are hard to solve for the appropriate expansion coefficients. The procedure presented here seems to be more desirable primarily for two reasons. One, we are able to set up the spectral representations of the angular densities in terms of normal modes directly by using G_0, and, two, the aforementioned restricted Fourier amplitudes possess a factorization property, similar to the one discussed for the planar case in Ref. 4, which allows one to decouple the appropriate integral equations for determing the unknown quantities. As in the Case–Zweifel approach, our procedure seems to work for homogeneous or inhomogeneous problems for regions whose boundary coincides with one of the coordinate surfaces for which the Helmholtz equation is separable. For the purposes of illustration we solve some spherically symmetric albedo and Milne problems for the interior of a sphere and the Milne problem for the exterior of a black sphere. However, the procedure is kept sufficiently general to deal with the problems which satisfy the separability criterion of the Helmholtz equation.

In passing we remark here that the spherically symmetric problems (for homogeneous isotropically scattering media) have been discussed in the literature.[5–7] In particular, Erdmann and Siewart[7] use the basis set of spherically symmetric normal modes, obtained by Davison[5] and later by Mitsis,[6] to expand the Green's function. Their procedure in the light of our formulation results in regular integral equations for expansion coefficients. We shall see later that such regular equations cannot be reduced to singular ones by an application of reduction operators. The reason, which we merely state here, is that the reduction operators do not commute with the spectral representations of the angular density.

1. GENERAL FORMULATION USING THE GREEN'S FUNCTION FOR A PURELY ABSORBING MEDIUM

The one-speed transport equation we wish to consider is

$$(1+\mathbf{\Omega}\cdot\nabla)\Psi(\mathbf{r},\Omega) = (C/4\pi)\int d\Omega' P(\mathbf{\Omega}\cdot\mathbf{\Omega}')\Psi(\mathbf{r},\Omega') + Q(\mathbf{r},\Omega), \quad (1.1)$$

where in the usual notation $\Psi(\mathbf{r},\Omega)$ is the neutron angular density (or specific intensity of radiation in a homogeneous grey atmosphere), $\mathbf{\Omega}$ is the direction of neutron propagation at the position vector $\mathbf{r}$, $P(\mathbf{\Omega}\cdot\mathbf{\Omega}')$ is the rotationally invariant scattering kernel, C is the average cross section for producing secondary neutrons (or albedo for single scattering), Q is some given source, and ∇ is the gradient operator.

Let us consider a time-reversed adjoint equation for a purely absorbing medium, i.e.,

$$(1-\mathbf{\Omega}\cdot\nabla)G_0(\mathbf{r},-\mathbf{\Omega};\mathbf{r}_0,-\mathbf{\Omega}_0) = \delta(\mathbf{r}-\mathbf{r}_0)\delta(\mathbf{\Omega}\cdot\mathbf{\Omega}_0). \quad (1.2)$$

In a conventional way we combine Eqs. (1.1) and (1.2) to obtain

$$\begin{aligned}\Psi(\mathbf{r},\Omega)\Theta(\mathbf{r}\epsilon V) &= -\int_S d\Omega' dS'\, \hat{n}(\mathbf{r}_s')\cdot\Omega'\Psi(\mathbf{r}_s',\Omega')G_0(\mathbf{r},\Omega;\mathbf{r}_s',\Omega') \\ &\quad + (C/4\pi)\int_V d^3r' d\Omega' d\Omega''\, P(\Omega''\cdot\Omega')\Psi(\mathbf{r}',\Omega') \\ &\quad \times G_0(\mathbf{r},\Omega;\mathbf{r}',\Omega'') + \int_V d^3r' d\Omega'\, Q(\mathbf{r}',\Omega') \\ &\quad \times G_0(\mathbf{r},\mathbf{\Omega};\mathbf{r}',\mathbf{\Omega}'),\end{aligned} \quad (1.3)$$

where V is the volume in which $\Psi(\mathbf{r},\Omega)$ is to be determined, S is the surface bounding V, $\hat{n}(\mathbf{r}_s)$ is a unit normal at S pointing away from V, $\mathbf{r}_s$ is a point on S, and

$$\Theta(\mathbf{r}\epsilon V) = \begin{cases} 1 & \text{when } \mathbf{r} \text{ lies in } V \\ 0 & \text{otherwise} \end{cases} \quad (1.4)$$

is the general form of the Heaviside step function. Finally, in arriving at Eq. (1.3), we used the reciprocity relation

$$G_0(\mathbf{r},-\Omega;\mathbf{r}_0,-\mathbf{\Omega}_0) = G_0(\mathbf{r}_0,\mathbf{\Omega}_0;\mathbf{r},\Omega). \quad (1.5)$$

Just for the sake of comparison we also present the integral representation of $\Psi(\mathbf{r},\mathbf{\Omega})$ as one would obtain by employing the time reversed adjoint equation for $C \neq 0$, i.e.,

$$\begin{aligned}(1-\mathbf{\Omega}\cdot\nabla)G(\mathbf{r},-\Omega;\mathbf{r}_0,-\Omega_0) &= (C/4\pi)\int d\Omega' P(\mathbf{\Omega}\cdot\mathbf{\Omega}') \\ &\quad \times G(\mathbf{r},-\mathbf{\Omega}';\mathbf{r}_0,-\mathbf{\Omega}_0) + \delta(\mathbf{r}-\mathbf{r}_0)\delta(\mathbf{\Omega}\cdot\mathbf{\Omega}_0).\end{aligned} \quad (1.6)$$

The integral for $\Psi(\mathbf{r},\mathbf{\Omega})$ is then given by

$$\begin{aligned}\Psi(\mathbf{r},\Omega)\Theta(\mathbf{r}\epsilon V) &= -\int_S d\Omega' dS'\, \hat{n}(\mathbf{r}_s')\cdot\Omega'\Psi(\mathbf{r}_s',\Omega')G(\mathbf{r},\Omega;\mathbf{r}_s',\Omega') \\ &\quad + \int_V d^3r' d\Omega'\, Q(\mathbf{r}',\Omega')G(\mathbf{r},\Omega;\mathbf{r}',\Omega'),\end{aligned} \quad (1.7)$$

where the reciprocity relation (1.5) also satisfied by G was used in arriving at this equation.

We remark that, in contrast to the integral representation (1.7), here it seems that Eq. (1.3) is an integral equation for $\Psi(\mathbf{r},\Omega)$, since the second term on the right hand side involves $\Psi(\mathbf{r},\Omega)$. Furthermore, it may seem that one must either know the surface distribution [in Eq. (1.3)] *a priori* or solve an additional integral equation obtained by letting $\mathbf{r}\to\mathbf{r}_s$. In other words, if G_0 is to be an infinite-medium Green's function, then Eq. (1.3), in comparison with Eq. (1.7), is apparently an overspecification of the problem. However, as we shall see presently, such is not the case. For that purpose we introduce the restricted Fourier amplitudes, mentioned previously, and show that representations (1.3) and (1.7) are actually equivalent.

Let us first take the Fourier transform of Eq. (1.2) and use the reciprocity relation (1.5) to obtain

$$G_0(\mathbf{r}, \Omega; \mathbf{r}_0, \Omega_0) = \delta(\Omega \cdot \Omega_0) \frac{1}{(2\pi)^3} \int d^3K \frac{e^{i\mathbf{K}\cdot(\mathbf{r}-\mathbf{r}_0)}}{1 + i\mathbf{K}\cdot\Omega}. \quad (1.8)$$

Substitution of G_0 in Eq. (1.3) then yields

$$\Psi(\mathbf{r}, \Omega)\Theta(\mathbf{r}\epsilon V) = \frac{1}{(2\pi)^3} \int d^3K \frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{1 + i\mathbf{K}\cdot\Omega} \times \Big[- \int_S ds'\, e^{-i\mathbf{K}\cdot\mathbf{r}_s'} \hat{n}(\mathbf{r}_s')\cdot\Omega\Psi(\mathbf{r}_s', \Omega) + \frac{C}{4\pi} \int_V d^3r' d\Omega'\, e^{-i\mathbf{K}\cdot\mathbf{r}'} P(\Omega\cdot\Omega')\Psi(\mathbf{r}', \Omega') + \int d^3r'\, e^{-i\mathbf{K}\cdot\mathbf{r}'} Q(\mathbf{r}', \Omega)\Big]. \quad (1.9)$$

Without loss of generality we approximate $P(\Omega, \Omega')$ by a degenerate kernel consisting of spherical harmonics as shown below:

$$P(\Omega\cdot\Omega') = 4\pi \sum_{l=0}^{N} \sum_{m=-l}^{l} b_l Y_{lm}^*(\Omega_K) Y_{lm}(\Omega_K'). \quad (1.10)$$

Here we have used the rotational invariance of $P(\Omega\cdot\Omega')$ and utilized the $\mathbf{K}$ vector as the vertical axis of a coordinate system so that

$$\Omega_K = (\hat{k}\cdot\Omega, \phi) \quad \text{and} \quad \Omega_K' = (\hat{k}\cdot\Omega', \phi'),$$

with ϕ and ϕ' being the azimuthal angles of Ω and Ω' in a plane perpendicular to the $\mathbf{K}$ axis, respectively. Putting the expansion (1.10) in Eq. (1.9), we obtain

$$\Psi(\mathbf{r}, \Omega)\Theta(\mathbf{r}\epsilon V) = \frac{1}{(2\pi)^3} \int d^3K \frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{1 + i\mathbf{K}\cdot\Omega} \Big(\int_V d^3r' e^{-i\mathbf{K}\cdot\mathbf{r}'} Q(\mathbf{r}', \Omega) - \int_S ds'\, e^{-i\mathbf{K}\cdot\mathbf{r}_s'} \hat{n}(\mathbf{r}_s')\cdot\Omega\Psi(\mathbf{r}_s', \Omega) + C\sum_{l,m}^{N} b_l Y_{lm}^*(\Omega_K) R_{lm}(\mathbf{K})\Big), \quad (1.11)$$

where

$$R_{lm}(\mathbf{K}) = \int_V d^3r' d\Omega'\, e^{-i\mathbf{K}\cdot\mathbf{r}'} Y_{lm}(\Omega_K')\Psi(\mathbf{r}', \Omega) \quad (1.12)$$

are the restricted Fourier amplitudes. More explicitly, $R_{lm}(\mathbf{K})$ are the volume restricted Fourier transforms of the spherical harmonic moments of the angular density. If we take the Fourier transform of Eq. (1.11) and take the harmonic moments of the resulting equation, then we find that the $R_{lm}(\mathbf{K})$ satisfy an algebraic equation which is given by

$$\sum_{l=|m|}^{N} R_{lm}(\mathbf{K}) A_m(l, l'; K) = \int d\Omega \frac{Y_{l'm}(\Omega)}{1 + i\mathbf{K}\cdot\Omega} \Big[\int d^3r'\, e^{-i\mathbf{K}\cdot\mathbf{r}'} Q(\mathbf{r}', \Omega) - \int dS'\, \hat{n}(\mathbf{r}_s')\cdot\Omega\Psi(\mathbf{r}_s', \Omega) e^{-i\mathbf{K}\cdot\mathbf{r}_s'}\Big], \quad (1.13)$$

where

$$A_m(l, l'; K) = \delta_{ll'} - Cb_l \int d\Omega \frac{Y_{lm}^*(\Omega_K) Y_{l'm}(\Omega_K)}{1 + i\mathbf{K}\cdot\Omega}. \quad (1.14)$$

When the determinant (the dispersion function)

$$\Lambda_m(K) = \det|A_m(l, l'; K)| \quad (1.15)$$

of the system of equations (1.13) is not zero, we have the unique solution

$$R_{lm}(\mathbf{K}) = \sum_{l'=|m|}^{N} \frac{d_m(^{l}_{l';K})}{\Lambda_m(K)} \times \int d\Omega \frac{Y_{l'm}(\Omega)}{1 + i\mathbf{K}\cdot\Omega} \Big[\int d^3r'\, e^{-i\mathbf{K}\cdot\mathbf{r}'} Q(\mathbf{r}', \Omega) - \int ds' e^{-i\mathbf{K}\cdot\mathbf{r}_s'} \hat{n}(\mathbf{r}_s')\cdot\Omega\Psi(\mathbf{r}_s', \Omega)\Big], \quad (1.16)$$

where $d_m(^{l}_{l';K})$ are minors associated with indices (l, l').

From Eq. (1.16) in conjunction with Eq. (1.11) it is now clear that once the surface angular density $\Psi(\mathbf{r}, \Omega)$ is known, then $R_{lm}(\mathbf{K})$ are known. In consequence $\Psi(\mathbf{r}, \Omega)$ is determined everywhere in V. In other words, Eq. (1.3) is not an overspecification of the boundary value problems since all we require is the determination of the surface angular density.

We now proceed to examine some generally useful properties of the restricted amplitudes R_{lm} in spherical geometry.

(1) In linear relations (1.13) among R_{lm}, if we regard $\mathbf{K}$ as a complex vector, i.e., $\mathbf{K} = (K, \hat{k})$ such that K can be a complex number, and in the spherical geometry note that

$$Q(\mathbf{r}, \Omega) = Q(r, \hat{r}\cdot\Omega, \text{azimuth angles of } \mathbf{r} \text{ and } \Omega)$$

and

$$\Psi(\mathbf{r}_s, \Omega) = \Psi(r_s, \hat{r}_s\cdot\Omega, \text{azimuth angles of } \mathbf{r}_s \text{ and } \Omega),$$

then R_{lm} are independent of $\hat{k}$, and, furthermore, under the reflection $K \to -K$,

$$R_{lm}(-K) = (-1)^{l+m} R_{lm}(K). \quad (1.17)$$

(2) The $R_{lm}(K)$ are sectionally holomorphic in the complex K plane with a branch cut for $K = -i\infty$ to $-i$ and i to $i\infty$ independent of the value of m, and have poles at the zeros of the determinant $\Lambda_m(K)$ [see Eq. (1.16)], the number of which is dependent on the value of m. In particular, if, for a given value of m, $\Lambda_m(K)$ is zero for real values of K, then $R_{lm}(\mathbf{K})$ are not unique and must be interpreted as generalized functions.

(3) The $R_{lm}(K)$ diverge exponentially as $K \to \infty$.

(4) If $\Psi(\mathbf{r}, \Omega)$ is spherically symmetric, i.e.,

$$\Psi(\mathbf{r}, \Omega) = \Psi(r, \hat{r}\cdot\Omega), \quad (1.18)$$

then

$$R_{lm}(K) = \delta_{m0} R_{l0}(K). \quad (1.19)$$

In the following section we shall list some more properties of the restricted amplitudes relevant to the spherically symmetric problems which we consider in some detail. Our objective will be to utilize those properties in casting the integral representation (1.13) into a normal mode expansion form and also to produce auxiliary singular integral equations for expansion coefficients by an appropriate application of reduction operators. The procedure we follow is essentially that discussed in Ref. 1, but somewhat more general so that it can be readily adapted to problems lacking azimuthal symmetry.

2. GENERAL FORMULATION FOR SPHERICALLY SYMMETRIC PROBLEMS

Under spherical symmetry Eq. (1.11), by virtue of property (4) (Sec. 1), becomes

$$\Psi(r,\mu)\Theta(\mathbf{r}\epsilon V) = \frac{1}{(2\pi)^3}\int d^3K \frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{1+i\mathbf{K}\cdot\mathbf{\Omega}}\left(q(K;\Omega) - F(K,\hat{k}\cdot\mathbf{\Omega}) + \frac{C}{4\pi}\sum_{l=0}^{N} b_l(2l+1)P_l(\hat{k}\cdot\mathbf{\Omega})R_l(K)\right), \tag{2.1}$$

where $\mu = \hat{r}\cdot\mathbf{\Omega}$,

$$q(K,\mathbf{\Omega}) = \int d^3r'\, e^{-i\mathbf{K}\cdot\mathbf{r}'}Q(\mathbf{r}',\mathbf{\Omega}), \tag{2.2}$$

$$F(K,\hat{k}\cdot\mathbf{\Omega}) = \epsilon\int_S ds'\, e^{-i\mathbf{K}\cdot\mathbf{r}'_s}\hat{n}(\mathbf{r}'_s)\cdot\mathbf{\Omega}\Psi(r_s,\hat{r}'_s\cdot\mathbf{\Omega}), \tag{2.3}$$

$$R_l(K) = \int_V d^3r'd\Omega'\, e^{-i\mathbf{K}\cdot\mathbf{r}'}P_l(\hat{k}\cdot\mathbf{\Omega}')\Psi(r',\mu'), \tag{2.4}$$

and

$$\Theta(\mathbf{r}\epsilon V) = \begin{cases}\Theta(r_s - r) & \text{for interior problems}\\ \Theta(r - r_s) & \text{for exterior problems.}\end{cases}$$

In (2.3) the unit normal $\hat{n}(\mathbf{r}'_s)$ at the surface of the sphere (of radius r_s) now always points in the direction of increasing r, whence $\epsilon = +1$ (-1) for interior (exterior) problems.

Assuming that if there is any internal or external source then it is independent of Ω and isotropic in the spatial coordinate, i.e.,

$$Q(\mathbf{r},\mathbf{\Omega}) = Q(r), \tag{2.5}$$

then we can reduce the set of algebraic equations (1.13) (for $m = 0$) to

$$\sum_{l=0}^{N} R_l(K)A(l,l';K) = 2\pi\int_{-1}^{1}dt\frac{P_{l'}(t)}{1+iKt}[q(K) - F(K,t)], \tag{2.6}$$

where now

$$A(l,l';K) = \delta_{ll'} - cb_l(l+\tfrac{1}{2})\int_{-1}^{1}d\mu\frac{P_l(\mu)P_{l'}(\mu)}{1+iK\mu}, \tag{2.7}$$

$$q(K) = \int_V d^3r'\, e^{-i\mathbf{K}\cdot\mathbf{r}'}Q(r'), \tag{2.8}$$

and

$$F(K,t) = \epsilon 2\pi r_s^2\int_{-1}^{1}d\mu'\mu' e^{-iKr_st\mu'} \times J_0(Kr_s[(1-\mu'^2)(1-t^2)]^{1/2})\Psi(r_s,\mu'). \tag{2.9}$$

In what follows our efforts will be directed toward the conversion of the Fourier integral representation (2.1) of $\Psi(r,\mu)$ into the spectral representation in which the functions occurring are the spherically symmetric normal modes of the transport equation and in which the integral is over the spectrum of the transport operator. For instance, for the interior problems $(r < r_s)$ such a representation would be of the form

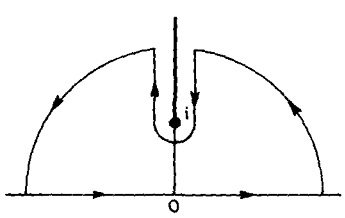

FIG. 1. Contour in the upper half of the complex K plane.

$$\Psi(r,\mu)\Theta(r_s - r) = \Psi_0(r,\mu) + \frac{1}{2\pi i}\int_0^1 d\nu\Gamma_<(\nu)E_\nu(r,\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)E_{\nu_j}(r,\mu) \tag{2.10}$$

where $\Psi_0(r,\mu)$ is the source angular density, $E_\nu(r,\mu)$ and $E_{\nu_j}(r,\mu)$ are the continuum and the discrete regular modes corresponding to continuum and discrete spectra of the transport operator, respectively, and $\Gamma_<(\nu)$ and $\Gamma_<^0(\nu_j)$ are the expansion coefficients which are to be determined by the boundary condition at $r = r_s$. As mentioned previously, since such integral equations are regular and hard to manage analytically, we shall produce auxiliary singular integrals for the same expansion coefficients by an application of reduction operators [see Eqs. (2.37), (2.38)] to Eq. (2.1). Thus, for interior problems such equations would be of the form

$$\Psi(r_s,\mu) = \Psi_R(r_s,\mu) + \frac{1}{2\pi i}\int_0^1 d\nu\Gamma_<(\nu)[L_\nu(\mu) + L_{-\nu}(\mu)] + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)[L_{\nu_j}(\mu) + L_{-\nu_j}(\mu)], \tag{2.11}$$

where $\Psi_R(r_s,\mu)$ is the transformed source angular density at the surface and $L_\nu(\mu)$, $L_{\nu_j}(\mu)$ [see Eqs. (4.14), (4.15)] are the reduced normal modes with a Cauchy type kernel. Analogous equations for exterior problems will also be given.

Now, in order to convert the Fourier integral representation (2.1) of $\Psi(r,\mu)$ into the spectral representations (for interior and exterior problems), our rather straightforward procedure (cf. Ref. 1) will consist of isolating the angular integrals over $\mathbf{K}$ and extending the integral with respect to K over the whole line $-\infty < K < \infty$. The latter step is possible by virtue of the reflection property (1.17). Then, by appropriately rearranging the integrands to insure convergence we shall consider the contour in the upper half complex K plane, as shown in Fig. 1, and apply the Cauchy formula. For that purpose we need to know the relevant singular and algebraic properties of the restricted amplitude $R_l(K)$ [in addition to the property (2), Sec. 1] explicitly which will be useful in obtaining the appropriate spectral representations of angular densities for interior and exterior properties.

Properties of $R_l(K)$

For the sake of convenience we separate out the source term in the linear relations (2.6) by writing

$$R_l(K) = \tilde{R}_l(K) + S_l(K) \tag{2.12}$$

so that $\tilde{R}_l(K)$ and $S_l(K)$ satisfy the following linear relations:

$$\sum_{l=0}^{N}\tilde{R}_l(K)A(l,l';K) = -2\pi\int_{-1}^{1}dt\frac{P_{l'}(t)}{1+iKt}F(K,t) \tag{2.13}$$

and

$$\sum_{l=0}^{N}S_l(K)A(l,l';K) = 2\pi q(K)\int_{-1}^{1}dt\frac{P_{l'}(t)}{1+iKt}. \tag{2.14}$$

For interior problems $(r < r_s)$ we further affect the decomposition of $\tilde{R}_l(K)$ into two parts,

$$2\tilde{R}_l(K) = R_l^{(1)}(K) + R_l^{(2)}(K), \tag{2.15}$$

such that $R_l^{(1),(2)}(K)$ converges in the upper (lower) half of the complex K plane. One may readily show that in (2.9) if we use the following expansions [8]:

$$e^{-iKr_s tt'}J_0(Kr_s[(1-t^2)(1-t'^2)])$$
$$= \sum_{n=0}^{\infty}(-1)^n i^n j_n(Kr_s)P_n(t)P_n(t') \quad (2.16)$$

and

$$2j_n(z) = h_n^{(1)}(z) + h_n^{(2)}(z), \quad (2.17)$$

where $j_n(z)$ and $h_n^{(1),(2)}(z)$ are spherical Bessel and Hankel functions, then, explicitly,

$$\sum_{l=0}^{N} R_l^{(1),(2)}(K)A(l,l';K)$$
$$= -2\pi r_s^2 \sum_{n=0}^{\infty}(-1)^n i^n(2n+1)h_n^{(1),(2)}(Kr_s)$$
$$\times \int_{-1}^{1} dt'\Psi(r_s,t')P_n(t')\int_{-1}^{1}dt\frac{P_{l'}(t)P_n(t)}{1+iKt}. \quad (2.18)$$

1. Recurrence Relations for $R_l^n, R_l^{(1),(2)}$ and S_l

In the sets of linear relations (2.13), (2.14), and (2.18) if we write the matrix elements $A(l,l',K)$ as

$$A(l,l';K) = \delta_{ll'}(1-cb_l) + cb_l(l+\tfrac{1}{2})\int_{-1}^{1}d\mu\frac{P_l(\mu)P_{l'}(\mu)iK\mu}{1+iK\mu} \quad (2.19)$$

and use the recurrence relation for Legendre polynomia, i.e.,

$$(2l+1)\mu P_l(\mu) = (l+1)P_{l+1}(\mu) + lP_{l-1}(\mu), \quad (2.20)$$

then the aplitudes $\tilde{R}_l(K), R_l^{(1),(2)}(K)$, and $S_l(K)$ are seen to obey the following recurrence relations:

$$(2l+1)(1-cb_l)\tilde{R}_l(K) + iK[(l+1)\tilde{R}_{l+1}(K) + l\tilde{R}_{l-1}(K)]$$
$$= -\epsilon(2l+1)4\pi^2 r_s^2 i^l(-1)^l j_l(Kr_s)$$
$$\times \int_{-1}^{1}dt\, tP_l(t)\Psi(r_s,t), \quad (2.21)$$

$$(2l+1)(1-cb_l)R_l^{(1),(2)}(K) + iK[(l+1)R_{l+1}^{(1),(2)}(K)$$
$$+ lR_{l-1}^{(1),(2)}(K)]$$
$$= -\epsilon(2l+1)4\pi^2 r_s^2(-1)^l i^l h_l^{(1),(2)}(Kr_s)$$
$$\times \int_{-1}^{1}dt\, tP_l(t)\Psi(r_s,t), \quad (2.22)$$

and

$$(2l+1)(1-cb_l)S_l(K) + iK[(l+1)S_{l+1}(K) + lS_{l-1}(K)]$$
$$= 4\pi\delta_{l0}q(K). \quad (2.23)$$

2. Factorization of $\tilde{R}_l, R_l^{(1),(2)}$, and S_l

If $\mathfrak{D}_l(z)$ denotes any of $\tilde{R}_l(z), R_l^{(1),(2)}, S_l(z)$, where $z = i/K$, then from recurrence relations (2.21)–(2.23) we conclude that

$$\mathfrak{D}_l(z) = g_l(z)\mathfrak{D}_0(z) + W_l(z), \quad (2.24)$$

where $g_l(z)$ are a complete set of orthogonal polynomia in z (orthogonal in the Steiljes sense)[9] which satisfy the following recurrence relation:

$$(2l+1)z(1-cb_l)g_l(z) = (l+1)g_{l+1}(z) + lg_{l-1}(z),$$
$$g_0 \equiv 1. \quad (2.25)$$

Also, depending upon what $\mathfrak{D}_l(z)$ denotes, $W_l(z)$ are also polynomia in z which satisfy the corresponding recurrence relations (2.21)–(2.23) with $W_0 \equiv 0$.

3. Factorization of $\mathfrak{D}_l(z)$ in the Spectrum of the Transport Operator

If ν is any point in the simple spectrum of the transport operator defined in the complex z plane (i.e., $z = i/K$) and if we denote by

$$D_l(\nu) \equiv \mathfrak{D}_l^+(\nu) - \mathfrak{D}_l^-(\nu), \quad \text{if } \nu \in [-1,1], \quad (2.26)$$

and

$$D_l(\nu) \equiv \lim_{z\to\nu}(z-\nu)\mathfrak{D}_l(z), \quad \text{if } \nu \text{ is a simple zero of } \Lambda(\nu) \text{ such that } \nu \notin [-1,1], \quad (2.27)$$

where +(–) represents the boundary value as z approaches the cut from the top (bottom) (see Fig. 2), then

$$D_l(\nu) = g_l(\nu)D_0(\nu). \quad (2.28)$$

4. Auxiliary Set of Linear Relations for $\tilde{R}_l(K)$ and $R_l^{(1)}(K)$

If $\{B(l,n;K)\}$ and $\{B^{(1),(2)}(l,n;K)\}$ are three sets of $N \times N$ matrices whose elements are given by

$$B(l,n;K) = i^l\int_{-1}^{1}d\mu\frac{P_n(\mu)}{1+iK\mu}\left((1-cb_l)(2l+1)\frac{P_l(\mu)}{j_l(Kr_s)}\right.$$
$$\left.-Kl\frac{P_{l-1}(\mu)}{j_{l-1}(Kr_s)} + K(l+1)\frac{P_{l+1}(\mu)}{j_{l+1}(Kr_s)}\right) \quad (2.29)$$

and

$$B^{(1),(2)}(l,n;K)$$
$$= i^l\int_{-1}^{1}d\mu\frac{P_n(\mu)}{1-iK\mu}\left((1-cb_l)(2l+1)\frac{P_l(\mu)}{h_l^{(1),(2)}(Kr_s)}\right.$$
$$\left.-Kl\frac{P_{l-1}(\mu)}{h_{l-1}^{(1),(2)}(Kr_s)} + K(l+1)\frac{P_{l+1}(\mu)}{h_{l+1}^{(1),(2)}(Kr_s)}\right), \quad (2.30)$$

respectively, then $\tilde{R}_l(K)$ and $R_l^{(1),(2)}(K)$ also satisfy the set of linear algebraic relations given by

$$\sum_{l=0}^{N}\tilde{R}_l(K)B(l,n;K) = -\epsilon 8\pi^2 r_s^2\int_{-1}^{1}d\mu\,\mu P_n(\mu)\frac{\Psi(r_s,\mu)}{1+iK\mu} \quad (2.31)$$

and

$$\sum_{l=0}^{N}R_l^{(1),(2)}(K)B^{(1),(2)}(l,n;K)$$
$$= -\epsilon 8\pi^2 r_s^2\int_{-1}^{1}d\mu\,\mu P_n(\mu)\frac{\Psi(r_s,\mu)}{1-iK\mu}. \quad (2.32)$$

[(1) with (1) and (2) with (2).] See Appendix A for proof.

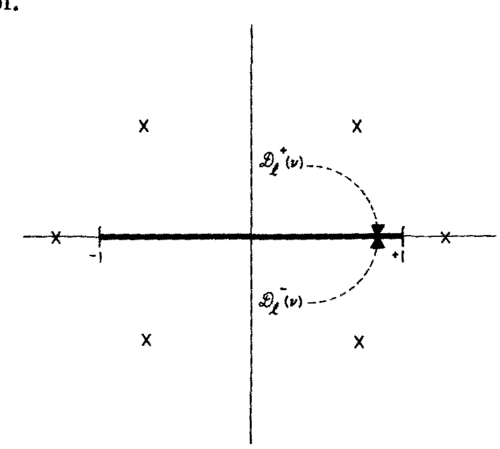

FIG. 2. Boundary values of $\mathfrak{D}_l(z)$ across the branch cut. — represents branch cut for $\mathfrak{D}_l(z)$; × represent poles of $\mathfrak{D}_l(z)$.

5. Reflection Properties
Under the reflection $K \to -K$,

$$\bar{R}_l(-K) = (-1)^l \bar{R}_l(K), \tag{2.33}$$

$$R_l^{(1)}(-K) = (-1)^l R_l^{(2)}(K), \tag{2.34}$$

$$S_l(-K) = (-1)^l S_l(K), \tag{2.35}$$

and

$$R_l(-K) = (-1)^l R_l(K). \tag{2.36}$$

We shall now utilize these properties in affecting the transition from the Fourier integral representation (2.1) of the angular density to the spectral representation of the form given by Eq. (2.10). In particular, we shall see that factorization property (3) stated by Eq. (2.28) will be useful in reducing the problem of solving several integral equations to that of solving only one and the property (4) will be of crucial importance in generating singular normal modes by means of reduction operators. Thus, let us introduce those operators here.

Reduction Operators

Consider two classes of operators $T_l^{(1)}$ and $T_l^{(2)}$ defined by

$$T_l^{(1)} \equiv \lim_{r \to r_s} \int_r^\infty dr'\, r'^2 h_l^{(2)}\left(\frac{ir'}{\mu}\right) \int_{-1}^1 d\mu' P_l(\mu') \times \left(1 + \mu' \frac{\partial}{\partial r'} + \frac{1-\mu'^2}{r'} \frac{\partial}{\partial \mu'}\right), \quad \mu < 0, \tag{2.37}$$

and

$$T_l^{(2)} \equiv \lim_{r \to r_s} \int_{-\infty}^r dr'\, r'^2 h_l^{(2)}\left(\frac{ir'}{\mu}\right) \int_{-1}^1 d\mu' P_l(\mu') \times \left(1 + \mu' \frac{\partial}{\partial r'} + \frac{1-\mu'^2}{r'} \frac{\partial}{\partial \mu'}\right), \quad \mu > 0. \tag{2.38}$$

We shall associate $T_l^{(1)}$ with interior and $T_l^{(2)}$ with exterior problems.

3. INTERIOR PROBLEMS

Consider now the particular situation involving determination of the angular density inside a sphere of radius r_s for a given incident angular density. Without loss of generality we let the internal source (if any) to be a point source at the origin, i.e.,

$$Q(r) = q\delta(r). \tag{3.1}$$

Equation (2.1) then becomes

$$\Psi(r,\mu)\Theta(r_s - r) = \frac{1}{(2\pi)^3} \int d^3K \frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{1 + i\mathbf{K}\cdot\mathbf{\Omega}} \left[q - F(K, \hat{k}\cdot\Omega) + \frac{C}{4\pi} \sum_{l=0}^N b_l(2l+1) P_l(\hat{k}\cdot\Omega) R_l(K)\right]. \tag{3.2}$$

Using the cosine formula for $\hat{k}\cdot\hat{r}$ in the exponential, i.e.,

$$\hat{k}\cdot\hat{r} = t\mu + [(1-t^2)(1-\mu^2)]^{1/2} \cos(\phi_K - \phi_r) \tag{3.3}$$

where $\mu = \hat{r}\cdot\mathbf{\Omega}$ and $t = \hat{k}\cdot\mathbf{\Omega}$, we perform the angular integrals with respect to K to obtain

$$\Psi(r,\mu)\Theta(r_s - r) = \frac{1}{4\pi^2} \int_0^\infty dK\, K^2 \int_{-1}^1 \frac{dt}{1+iKt} e^{iKrt\mu} \times J_0(Kr[(1-t^2)(1-\mu^2)]^{1/2}) \times \left(q - F(K,t) + \frac{C}{4\pi} \sum_{l=0}^N b_l(2l+1) P_l(t) R_l(K)\right). \tag{3.4}$$

Under the reflection property (2.36) of $R_l(K)$ we may extend the integral with respect to K over the whole line and use the decompositions (2.12) and (2.15) of $R_l(K)$ and the reflection properties (2.34) and (2.35) of its components to re-express Eq. (3.4) as given below:

$$\Psi(r,\mu)\Theta(r_s - r) = \Psi_0(r,\mu) - \frac{1}{4\pi^2} \int_{-\infty}^\infty dK\, K^2 \times \int_{-1}^1 \frac{dt}{1+iKt} e^{iKrt\mu} J_0(Kr[(1-t^2)(1-\mu^2)]^{1/2}) \times \left(H^{(1)}(K,t) - \frac{C}{4\pi} \sum_{l=0}^N b_l(2l+1) P_l(t) R_l^{(1)}(K)\right), \tag{3.5}$$

where $\Psi_0(r,\mu)$, the source angular density, is defined by

$$\Psi_0(r,\mu) = \frac{q}{8\pi^2} \int_{-\infty}^\infty dK\, K^2 \int_{-1}^1 \frac{dt}{1+iKt} \times \sum_{n=0}^\infty i^n (2n+1) h_n^{(1)}(Kr) P_n(t) P_n(\mu) \times \left(1 - \frac{C}{4\pi} \sum_{l=0}^N b_l(2l+1) P_l(t) S_l(K)\right) \tag{3.6}$$

and

$$H^{(1)}(K,t) = \pi r_s^2 \sum_{l=0}^N (-1)^l i^l (2l+1) h_l^{(1)}(Kr_s) P_l(t) \times \int_{-1}^1 d\mu\, \mu P_l(\mu) \Psi(r_s,\mu). \tag{3.7}$$

Now consider the contour in the upper half of the complex K plane as shown in Fig. 1. Since the quantities $H^{(1)}$ and $R_l^{(1)}$ in (3.5) approach zero as $|K| \to \infty$ dominantly over the divergent terms involving the exponential and the Bessel functions, the contribution from the semicircle is zero. We also note that $H^{(1)}$ is regular in that part of the complex K plane. Thus, in view of the Plemelj's formula across the cut

$$1/(\nu - \mu)_\pm = [\mathcal{P}/(\nu - \mu)] \mp i\pi\delta(\nu - \mu), \tag{3.8}$$

where $\mathcal{P}$ is the Cauchy principle value, Eq. (3.5) becomes

$$\Psi(r,\mu)\Theta(r_s - r) = \Psi_0(r,\mu) + \frac{1}{4\pi^2 i} \int_0^1 \frac{d\nu}{\nu^4} \times \int_{-1}^1 dt\, e^{-rt\mu/\nu} J_0\left(\frac{ir}{\nu}\sqrt{(1-t^2)(1-\mu^2)}\right) \times \left[-z\pi i\,\nu H^{(1)}\left(\frac{i}{\nu},\nu\right)\delta(\nu - t) - \frac{C}{4\pi} \sum_{l=0}^N b_l(2l+1) \times P_l(t)\,[R_l^{(1)+}(\nu) - R_l^{(1)-}(\nu)]\, \mathcal{P}\frac{\nu}{\nu - t} + \frac{iC}{4} \times \sum_{l=0}^N b_l(2l+1)\nu P_l(\nu)\,[R_l^{(1)+}(\nu) + R_l^{(1)-}(\nu)]\,\delta(\nu - t)\right] + \frac{1}{2\pi i} \sum_{j=0}^M \int_{-1}^1 \frac{d t \nu_j}{\nu_j - t} e^{-rt\mu/\nu_j} J_0\left(\frac{ir}{\nu_j}[(1-t^2)(1-\mu^2)]\right) \times \frac{C}{4\pi} \sum_{l=0}^N b_l(2l+1) P_l(t) \lim_{z \to \nu_j} (z - \nu_j) R_l^{(1)}(z), \tag{3.9}$$

where we have put $z = i/K$ which means that $R_l^{(1)\pm}(\nu)$ are the boundary values of $R_l^{(1)}(z)$ in the complex z plane (see Fig. 2) as z approaches the cut from the top (bottom), ν_j are the nondegenerate roots of

$$\Lambda(\nu_j) \equiv \det |A(l, l'; \nu_j)| = 0, \quad j = 0, 1, \ldots, M, \quad (3.10)$$

and $M + 1$ is the total number of such roots.

Now Eq. (3.9) will give us an appropriate representation provided we can relate $R_l^{(1)+}(\nu) + R_l^{(1)-}(\nu)$ (the functionals occurring in the third integrand) to $R_l^{(1)+}(\nu) - R_l^{(1)}(\nu)$. This is simply because the latter quantities factorize in accordance with the formula (2.28) while the former ones do not; a property which is essential to reducing the number of unknown quantities we must compute by reducing the number of integral equations one needs to solve. Furthermore, it is also necessary to eliminate $H^{(1)}(i/\nu, \nu)$, because by definition (3.7), it contains the complete angular integral over the surface distribution, which is an unknown quantity. Both these requirements are met if we consider the difference of boundary values of the set of algebraic relations (2.18) among $R_l^{(\nu)}(K)$. Thus, in the z plane we have

$$\frac{iC}{4} \sum_{l=0}^{N} b_l(2l+1)\nu P_l(\nu)[R_l^{(1)+}(\nu) + R_l^{(1)-}(\nu)]$$
$$= 2\pi i\nu\, H^{(1)}\left(\frac{i}{\nu}, \nu\right) - \frac{1}{4\pi P_{l'}(\nu)} \sum_{l=0}^{N} [R_l^{(1)+}(\nu) - R_l^{(1)-}(\nu)]$$
$$\times [A^+(l, l'; \nu) + A^-(l, l^1; \nu)]. \quad (3.11)$$

If we now appeal to the factorization property (2.28) and use the relation (3.11) in Eq. (3.9), we obtain

$$\Psi(r,\mu)\Theta(r_s - r) = \Psi_0(r,\mu) + \frac{1}{2\pi i}\int_0^1 d\nu\, \Gamma_<(\nu) E_\nu(r,\mu)$$
$$+ \frac{1}{2\pi i} \sum_{j=0}^{M} \Gamma_<^0(\nu_j) E_{\nu_j}(r,\mu), \quad (3.12)$$

where

$$\Gamma_<(\nu) = -[R_0^+(\nu) - R_0^-(\nu)]/8\pi^2\nu^4 \quad (3.13)$$

$$\Gamma_<^0(\nu_j) = \lim_{z\to\nu_j} (z - \nu_j) R_0(z)/4\pi^2\nu_j^2, \quad j = 0, 1, \ldots, M, \quad (3.14)$$

$$E_\nu(r,\mu) = C \sum_{l=0}^{N} b_l(2l+1) g_l(\nu) \int_{-1}^{1} dt\, P_l(t)\, e^{-rt\mu/\nu}$$
$$\times J_0\left(\frac{ir}{\nu}[(1-t^2)(1-\mu^2)]^{1/2}\right)$$
$$\times \left(\mathcal{P}\frac{\nu}{\nu - t} + \pi i\nu\delta(\nu - t)\frac{A^+(l,0;\nu) + A^-(l,0;\nu)}{A^+(l,0;\nu) - A^-(l,0;\nu)}\right), \quad (3.15)$$

and

$$E_{\nu_j}(r,\mu) = C \sum_{l=0}^{N} b_l(2l+1)\, g_l(\nu_j) \int_{-1}^{1} dt \frac{\nu_j}{\nu_j - t}$$
$$\times P_l(t)\, e^{-rt\mu/\nu_j} J_0\left(\frac{ir}{\nu_j}[(1-t^2)(1-\mu^2)]^{1/2}\right). \quad (3.16)$$

Here $E_\nu(r,\mu)$ and $E_{\nu_j}(r,\mu)$ are the continuum and the discrete spherically symmetric regular normal modes of the transport equation and $\Gamma_<(\nu)$ and $\Gamma_<^0(\nu_j)$ are the corresponding expansion coefficients which are to be determined from the given boundary condition at $r = r_s$. An equation which determines these coefficients is obtained by letting $r \to r_s$ in Eq. (3.12). Thus,

$$\Psi(r_s,\mu) = \Psi_0(r_s,\mu) + \frac{1}{2\pi i}\int_0^1 d\nu\, \Gamma_<(\nu) E_\nu(r_s \mu)$$
$$+ \frac{1}{2\pi i}\sum_{j=0}^{M} \Gamma_<^0(\nu_j) E_{\nu_j}(r_s,\mu), \ \mu < 0. \quad (3.17)$$

This is a regular integral equation which clearly lacks the appropriate Cauchy type kernel. In consequence it is not manageable by the standard techniques despite the fact that E_ν and E_{ν_j} form a complete set. To prove the completeness of these modes, we shall resort to the application of reduction operator $T_l^{(1)}$ to Eq. (3.4) and produce an auxiliary soluble singular integral equation which will contain the same coefficients, defined by Eqs. (3.13) and (3.14), occurring in Eq. (3.17). It is worthwhile to point out here that a direct application of $T_l^{(1)}$ to Eq. (3.12) is not possible because it does not commute with the spectral representation. In fact one may very quickly convince oneself that such a direct application will result in the divergence of the right hand side of Eq. (3.12).

Before we consider the above step, let us write down the spectral representation of the source angular density $\Psi_0(r,\mu)$ [see Eq. (3.6)]. The same procedure we used above yields

$$\Psi_0(r,\mu) = -\frac{1}{2\pi i}\int_0^1 d\nu \frac{S_0^+(\nu) - S_0^-(\nu)}{8\pi^2\nu^4} E_\nu^{(1)}(r,\mu)$$
$$-\frac{1}{2\pi i}\sum_{j=0}^{M} E_{\nu_j}^{(1)}(r,\mu) \lim_{z\to\nu_j} (z - \nu_j) S_0(z), \quad (3.18)$$

where

$$E_\nu^{(1)}(r,\mu) = C \sum_{l=0}^{N} b_l(2l+1) g_l(\nu)$$
$$\times \int_{-1}^{1} dt\, P_l(t) \sum_{n=0}^{\infty} i^n(2n+1) h_n^{(1)}\left(\frac{ir}{\nu}\right) P_n(t) P_n(\mu)$$
$$\times \left(\mathcal{P}\frac{\nu}{\nu - \mu} + i\pi\nu\delta(\nu - t)\frac{A^+(l,0;\nu) + A^-(l,0;\nu)}{A^+(l,0;\nu) - A^-(l,0;\nu)}\right) \quad (3.19)$$

and

$$E_{\nu_j}^{(1)}(r,\mu) = C\sum_{l=0}^{N} b_l(2l+1) g_l(\nu_j) \int_{-1}^{1} dt\, P_l(t) \frac{\nu_j}{\nu_j - t}$$
$$\times \sum_{n=0}^{\infty} i^n(2n+1)\, h_n^{(1)}\left(\frac{ir}{\nu_j}\right) P_n(t) P_n(\mu). \quad (3.20)$$

The explicit values of the amplitudes $[S_0^+(\nu) - S_0^-(\nu)]$ and $\lim_{z\to\nu_j}(z - \nu_j)S_0(z)$ are given in Appendix B.

4. APPLICATION OF THE REDUCTION OPERATOR TO INTERIOR PROBLEMS

One may readily show that the result of an application of the reduction operator $T_l^{(1)}$, defined by Eq. (2.37), to Eq. (3.4) is

$$\lim_{r\to r_s} \int_{-\infty}^{\infty} dK\, K^2 R_l(K) \int_r^{\infty} dr'\, r'^2 h_l^{(2)}\left(\frac{ir'}{\mu}\right) j_l(Kr') = 0,$$
$$\mu > 0. \quad (4.1)$$

In terms of $R_l^{(1)}(K)$ and $S_l(K)$ [see decomposition equations (2.12), (2.15) and reflection properties (2.33)–(2.36)], we may rewrite Eq. (4.1) as

$$\lim_{r\to r_s} \int_{-\infty}^{\infty} dK\, K^2 [R_l^{(1)}(K) + S_l(K)] \int_r^s dr'\, r'^2 h_l^{(2)}\left(\frac{ir'}{\mu}\right)$$
$$\times j_l(Kr') = 0, \quad \mu < 0. \quad (4.2)$$

Now the integral over the spherical Hankel and Bessel functions has the interesting replication type property[8] that

$$\int_r^{\infty} dr'\, r'^2 h_l^{(2)}\left(\frac{ir'}{\mu}\right) j_l(Kr') = \frac{r^2\mu}{2i(1 + K^2\mu^2)}$$

$$\times\left[iK\mu h_l^2\left(\frac{ir}{\mu}\right)j_{l+1}(Kr)+h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)j_l(Kr)\right]. \quad (4.3)$$

If we put this in Eq. (4.2), we obtain

$$\tfrac{1}{2}\lim_{r\to r_s}\int_{-\infty}^{\infty}dK\,K^2\frac{R_l^{(1)}(K)}{1+K^2\mu^2}\left[iK\mu h_l^{(2)}\left(\frac{ir}{\mu}\right)h_{l+1}^{(1)}(Kr)+h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(1)}(Kr)+iK\mu h_l^{(2)}\left(\frac{ir}{\mu}\right)h_{l+1}^{(2)}(Kr)+h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(2)}(Kr)\right]$$
$$+\lim_{r\to r_s}\int_{-\infty}^{\infty}dK\,K^2\frac{S_l(K)}{1+K^2\mu^2}\left[iK\mu h_l^{(2)}\left(\frac{ir}{\mu}\right)\times h_{l+1}^{(1)}(Kr)+h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(1)}(Kr)\right]=0,\quad \mu=0, \quad (4.4)$$

where we have expressed the spherical Bessel functions as sums of two kinds of spherical Hankel functions and used the reflection property (2.35) of $S_l(K)$. We note that the denominator in both integrands is the product of two Cauchy type kernels, viz., $(l-iK\mu)$ and $(l+iK\mu)$ and the numerators associated with $R_l^{(1)}(K)$ and $S_l(K)$ strongly suggest an application of the Christoffel-Darboux formula.[10] In other words we can write

$$iK\mu h_l^{(2)}\left(\frac{ir}{\mu}\right)h_{l+1}^{(1)}(Kr)+h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(1)}(Kr)$$
$$=(1-iK\mu)\left[h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(1)}(Kr)+\frac{(-1)^{l+1}\mu}{ir}\times\sum_{m=0}^{l}(-1)^m(2m+1)h_m^{(2)}\left(\frac{ir}{\mu}\right)h_m^{(1)}(Kr)\right] \quad (4.5)$$

and

$$iK\mu h_l^{(2)}\left(\frac{ir}{\mu}\right)h_{l+1}^{(2)}(Kr)+h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(2)}(Kr)$$
$$=(1+iK\mu)\left[h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(2)}(Kr)-\frac{\mu}{ir}\sum_{m=0}^{l}(2m+1)\times h_m^{(2)}\left(\frac{ir}{\mu}\right)h_m^{(2)}(Kr)\right]. \quad (4.6)$$

These identities, by the Christoffel–Darboux theorem, are in fact common to all orthogonal functions. Here they permit us to rewrite (4.4) such that each integral contains only one Cauchy kernel instead of a product of two. In consequence Eq. (4.4) becomes

$$\tfrac{1}{2}\lim_{r\to r_s}\int_{-\infty}^{\infty}dK\,K^2R_l^{(1)}(K)\left(\frac{U_l^{(1)}(K;r,\mu)}{1+iK\mu}+\frac{U_l^{(2)}(K;r,\mu)}{1-iK\mu}\right)$$
$$+\lim_{r\to r_s}\int_{-\infty}^{\infty}dK\,K^2S_l(K)\frac{U_l^{(1)}(K;r,\mu)}{1+iK\mu}=0,\quad \mu<0, \quad (4.7)$$

where

$$U^{(1)}(K;r,\mu)=h_{l+1}^{(2)}\left(\frac{ir}{\mu}\right)h_l^{(1)}(Kr)+(-1)^{l+1}\frac{\mu}{ir}\times\sum_{m=0}^{l}(-1)^m(2m+1)h_m^{(2)}\left(\frac{ir}{\mu}\right)h_m^{(1)}(Kr) \quad (4.8)$$

and

$$U^{(2)}(K;r,\mu)=(-1)^lU^{(1)}(-K;r,\mu). \quad (4.9)$$

As before, we now consider the same contour, as shown in Fig. 1, and apply the Cauchy theorem to Eq. (4.7). Noting that $R_l^{(1)}(K)$ in the first integral and $U_l^{(1)}(K;r,\mu)$ in the second integral converge dominantly in the upper half complex K plane over their respective associates, we have zero contribution from the semicircle. In consequence, we obtain

$$\int_0^1\frac{d\nu}{\nu^4}\frac{\nu U_l^{(1)}(i/\nu;r_s,\mu)}{\nu-\mu}\left[R_l^{(1)+}(\nu)-R_l^{(1)}(\nu)\right]$$
$$+\mathcal{P}\int_0^1\frac{d\nu}{\nu^4}\frac{\nu U_l^{(2)}(i/\nu;r_s,\mu)}{\nu+\mu}\left[R_l^{(1)+}(\nu)-R_l^{(1)-}(\nu)\right]$$
$$+\frac{i\pi}{\mu^3}U_l^{(2)}(-i/\mu;r_s,\mu)\left[R_l^{(1)+}(-\mu)+R_l^{(1)-}(-\mu)\right]$$
$$+2\pi\sum_{j=0}^{M}\frac{1}{\nu_j^2}\left[\frac{\nu_j}{\nu_j-\mu}U_l^{(1)}\left(\frac{i}{\nu_j};r_s,\mu\right)+\frac{\nu_j}{\nu_j+\mu}\times U_l^{(2)}\left(\frac{i}{\nu_j};r_s,\mu\right)\right]\lim_{z\to\nu_j}(z-\nu_j)R_l^{(1)}(z)$$
$$+2\int_0^1\frac{d\nu}{\nu^4}\frac{\nu U_l^{(1)}(i/\nu;r_s,\mu)}{\nu-\mu}\left[S_l^{(1)+}(\nu)-S_l^{(1)-}(\nu)\right]$$
$$+4\pi\sum_{j=0}^{M}\frac{1}{\nu_j^2}\frac{\nu_jU_l^{(1)}(i/\nu_j;r_s,\mu)}{\nu_j-\mu}\lim_{z\to\nu_j}(z-\nu_j)S_l(z)=0,$$
$$\mu<0. \quad (4.10)$$

In relating the sum $R_l^{(1)+}(-\mu)+R_l^{(1)-}(-\mu)$ of the boundary values of the amplitudes to their differences we slightly deviate from the previous procedure. Thus, instead of the linear relation (2.18) we now consider the auxiliary set (2.32) and take its difference of boundary values across the cut in the $z=i/K$ plane. The result for $n=0$ is

$$\sum_{l=0}^{N}\left[R_l^{(1)+}(-\mu)+R_l^{(1)-}(-\mu)\right]\times\left[B^{(1)+}(l,0;-\mu)-B^{(1)-}(l,0;-\mu)\right]$$
$$=-\sum_{l=0}^{N}\left[R_l^{(1)+}(-\mu)-R_l^{(1)-}(-\mu)\right]\left[B^{(1)+}(l,0;-\mu)+B^{(1)-}(l,0;-\mu)\right]-32\pi^3ir_s^2\mu^2\Psi(r_s,\mu). \quad (4.11)$$

The reason for introducing the auxiliary set of equations (2.32) should now be obvious, for, in order to incorporate the boundary condition in the reduced integral equation (4.10), we need to isolate the surface distribution $\Psi(r_s,\mu)$. Clearly the set (2.32) satisfies that requirement while the set (2.18) does not. Thus, in Eq. (4.10) if we note that

$$U_l^{(2)}(-i/\mu;r_s,\mu)=(-1)^l\mu/r_s \quad (4.12)$$

and multiply both sides by

$$(-1)^l\left[B^{(1)+}(l,0;\mu)-B^{(1)-}(l,0;\mu)\right],$$

sum over l from 0 to N, and use the identity (4.11), we get the following result after the appropriate application of the factorization formula (2.28) and rearrangement of terms:

$$\frac{1}{2\pi i}\int_0^1d\nu\,\Gamma_<(\nu)\left[L_\nu(\mu)+L_{-\nu}(\mu)\right]+\frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)\times\left[L_{\nu_j}(\mu)+L_{-\nu_j}(\mu)\right]=\Psi(r_s,\mu)-\Psi_R(r_s,\mu), \quad (4.13)$$

where $\Gamma_<(\nu)$ and $\Gamma_<^0(\nu_j)$ are the same coefficients as those occurring in the integral equation (3.17). They are defined by Eqs. (3.13) and (3.14), respectively, while the functions $L_\nu(\mu)$, $L_{\nu_j}(\mu)$, and $\Psi_R(r_s,\mu)$ are given by

$$L_\nu(\mu) = \frac{1}{2\pi i r_s} \sum_{l=0}^{N} (-1)^l U_l^{(2)}\left(\frac{i}{\nu}; r_s, \mu\right) g_l(\nu) \times \left([B^{(1)+}(l,0;-\mu) - B^{(1)-}(l,0;-\mu)]\,\mathcal{P}\frac{\nu}{\nu+\mu} + \pi i [B^{(1)+}(l,0;-\mu) + B^{(1)-}(l,0;-\mu)]\nu\delta(\nu+\mu)\right), \tag{4.14}$$

$$L_{\nu_j}(\mu) = \frac{1}{2\pi i r_s} \sum_{l=0}^{N} (-1)^l U_l^{(2)}\left(\frac{i}{\nu_j}; r_s, \mu\right) g_l(\nu_j) \times [B^{(1)+}(l,0;-\mu) - B^{(1)-}(l,0;-\mu)]\frac{\nu_j}{\nu_j+\mu}, \tag{4.15}$$

and

$$\Psi_R(r_s, \mu) = \frac{1}{8\pi^3 i}\int_0^1 \frac{d\nu}{\nu^4} L_{-\nu}(\mu)[S_0^+(\nu) - S_0^-(\nu)] + \frac{1}{4\pi^2 i}\sum_{j=0}^{M}\frac{L_{\nu_j}(\mu)}{\nu_j^2}\lim_{z\to\nu_j}(z-\nu_j)S_0(z). \tag{4.16}$$

Here Ψ_R is the transformed angular density at the surface of the sphere due to the source. We note that it is appropriately expressed in terms of the reduced normal modes $L_{-\nu}(\mu)$ and $L_{\nu_j}(\mu)$ and the continuum and discrete amplitudes $S_0^+(\nu) - S_0^-(\nu)$ and $\lim_{z\to\nu_j}(z-\nu_j)S_0(z)$. The explicit values of these amplitudes are given in the Appendix B [see Eqs. (B5) and (B7)]. We remark here that in the integral equation (4.13) the reduced normal modes $L_\nu(\mu)$ are singular for $\nu > 0, \mu < 0$ while they are regular for $\nu > 0, \mu > 0$. Furthermore, from the structure of these modes it is clear that Eq. (4.13) is a singular integral equation with a Cauchy type kernel which can be solved by the standard technique due to Muskhelishvili.[11] We consider the applications of this singular integral equation next.

Applications (Interior Problems)

First we shall present the most general solution of Eq. (4.13) and then consider the particular problems such as the Milne and the albedo problems for the interior of the sphere.

Let us at the outset separate out the Fredholm term in Eq. (4.13) as follows. For any point z in the complex plane define

$$Y(\nu; z) = \frac{1}{2\pi i r_s}\sum_{l=0}^{N}(-1)^l U_l^{(2)}\left(\frac{i}{\nu}; r_s, \mu\right) g_l(\nu) B^{(1)}(l,0;z). \tag{4.17}$$

Then

$$L_\nu(\mu) = [Y^+(\nu;-\mu) - Y^-(\nu;-\mu)]\mathcal{P}[\nu/(\nu+\mu)] + \pi i [Y^+(-\mu;-\mu) + Y^-(-\mu;-\mu)]\nu\delta(\nu+\mu) \tag{4.18}$$

or

$$L_\nu(\mu) = [Y^+(\nu;-\mu) - Y^-(\nu;-\mu) + Y^+(-\mu;-\mu) - Y^-(-\mu;-\mu)][\nu/(\nu+\mu)] - [Y^+(-\mu;-\mu) - Y^-(-\mu;-\mu)]\mathcal{P}[\nu/(\nu+\mu)] + \pi i [Y^+(-\mu;-\mu) + Y^-(-\mu;-\mu)]\nu\delta(\nu+\mu). \tag{4.19}$$

One may readily check that in Eq. (4.19) the first term on the right-hand side enclosed in the square bracket is proportional to a polynomial in ν, i.e., it is degenerate. Hence, if we substitute $L_\nu(\mu)$ in Eq. (4.13) by Eq. (4.19) and separate out the Fredholm term, we get

$$[Y^+(-\mu;-\mu) - Y^-(-\mu;-\mu)]\mathcal{P}\frac{1}{2\pi i}\int_0^1 d\nu \frac{\nu\Gamma_<(\nu)}{\nu+\mu} + \tfrac{1}{2}[Y^+(-\mu;-\mu) + Y^-(-\mu;-\mu)]\mu\Gamma_<(-\mu) = \frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_<(\nu)L_{-\nu}(\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j) \times [L_{\nu_j}(\mu) + L_{-\nu_j}(\mu)] - \Psi(r_s;\mu) + \Psi_R(r_s;\mu) + C(\mu), \quad \mu < 0, \tag{4.20}$$

where

$$C(\mu) = \frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_<(\nu)\frac{\nu}{\nu+\mu}[Y^+(\nu;-\mu) - Y^-(\nu;-\mu) + Y^+(-\mu;-\mu) - Y^-(-\mu;-\mu)] \tag{4.21}$$

is the degenerate Fredholm term.

The most general solution of Eq. (4.20) is[9]

$$\nu\Gamma_<(\nu) = -\frac{1}{2\pi i}\int_0^1 d\nu' \frac{X^-(\nu')}{Y^-(\nu';\nu')} I_{\nu'}(\nu)\tilde{\Psi}(-\nu'), \quad \nu > 0, \tag{4.22}$$

where

$$\tilde{\Psi}(\mu) = \frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_<(\nu)L_{-\nu}(\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)[L_{\nu_j}(\mu) + L_{-\nu_j}(\mu)] - \Psi(r_s,\mu) + \Psi_R(r_s,\mu) + C(\mu), \quad \mu < 0, \tag{4.23}$$

$$I_{\nu'}(\nu) = \left(\frac{1}{X^-(\nu)} - \frac{1}{X^+(\nu)}\right)\mathcal{P}\frac{1}{\nu'-\nu} + i\pi\left(\frac{1}{X^+(\nu)} + \frac{1}{X^-(\nu)}\right)\delta(\nu'-\nu), \tag{4.24}$$

and

$$X(z) = \frac{1}{(1-z)^{M+1}}\exp\left(\frac{1}{2\pi i}\int_0^1 \frac{d\nu}{\nu - z}\ln\frac{Y^+(\nu;\nu)}{Y^-(\nu;\nu)}\right). \tag{4.25}$$

The additional conditions that determine $\Gamma_<^0(\nu_j)$ are

$$\int_0^1 d\nu\,\nu^l \frac{X^-(\nu)}{Y^-(\nu;\nu)}\psi(-\nu) = 0, \quad l = 0,1,\ldots,M, \tag{4.26}$$

where, as a reminder, $M+1$ is the total number of zeros of the characteristic equation (3.10), i.e.,

$$\Lambda(\nu_j) \equiv \det|A(l,l';\nu_j)| = 0, \quad j = 0,1,\ldots,M,$$

or more conveniently

$$\sum_{l=0}^{N} g_l(\nu_j)A(l,0;\nu_j) = 0, \quad j = 0,1,\ldots,M. \tag{4.27}$$

We note that the set of equations (4.26) will give us precisely $M+1$ number of equations, from which all $\Gamma_<^0(\nu_j)$ can be determined.

1. Milne Problem

The boundary condition for the Milne problem is

$$\Psi(r_s,\mu) = 0, \quad \mu < 0, \tag{4.28}$$

while the angular density due to the source is given by $\Psi_0(r,\mu)$ [see Eq. (3.18)], which under the reduction operator is transformed to $\Psi_R(r_s,\mu)$ [see Eq. (4.16)]. In consequence, $\tilde{\psi}(\mu)$, as defined by Eq. (4.23), becomes

$$\tilde{\psi}(\mu) = \frac{1}{2\pi i}\int_0^1 d\nu\, \Gamma_<(\nu) L_{-\nu}(\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)[L_{\nu_j}(\mu) + L_{-\nu_j}(\mu)] + \Psi_R(r_s,\mu) + C(\mu), \quad \mu < 0. \tag{4.29}$$

Putting this in Eq. (4.22), we get

$$-\nu\Gamma_<(\nu) = \frac{1}{2\pi i}\int_0^1 d\nu_1 \Gamma_<(\nu_1)\int_0^1 d\nu' \frac{X^-(\nu')}{Y^-(\nu';\nu')} \times I_{\nu'}(\nu) L_{-\nu_1}(-\nu') + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)\frac{1}{2\pi i}\int_0^1 d\nu' \times \frac{X^-(\nu')}{Y^-(\nu';\nu)} I_{\nu'}(\nu)[L_{\nu_j}(-\nu') + L_{-\nu_j}(-\nu')] + \frac{1}{2\pi i}\int_0^1 d\nu' \frac{X^-(\nu')}{Y^-(\nu';\nu')} I_{\nu'}(\nu)[\Psi_R(r_s;-\nu') + C(-\nu')]. \tag{4.30}$$

Equations which determine $\Gamma_<^0(\nu_j)$ are

$$\int_0^1 d\nu\,\Gamma_<(\nu)\int_0^1 d\nu' \nu'^{l} \frac{X^-(\nu')}{Y^-(\nu';\nu')} L_{-\nu}(-\nu') + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j) \times \int_0^1 d\nu'\nu'^{l} \frac{X^-(\nu')}{Y^-(\nu';\nu')}[L_{\nu_j}(-\nu') + L_{-\nu_j}(-\nu')] + \int_0^1 d\nu'\, \nu'^{l} \frac{X^-(\nu')}{Y^-(\nu';\nu')}[\Psi_R(r_s;-\nu') + C(-\nu')] = 0,$$
$$l = 0, 1, \ldots, M. \tag{4.31}$$

Equations (4.30) and (4.31) are ideally suited for asymptotic expansion for large r_s. In that case the first term on the right-hand side of Eq. (4.30) is exponentially small. We can then solve this equation in conjunction with Eq. (4.31) by iteration.

2. *Albedo Problem*

For the albedo problem

$$\Psi(r_s,\mu) = \delta(\mu - \mu_1), \quad \mu < 0,\ \mu_1 < 0, \tag{4.32}$$

and

$$\Psi_0(r_s,\mu) = \Psi_R(r_s,\mu) = 0. \tag{4.33}$$

Under these conditions

$$\tilde{\Psi}(\mu) = \frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_<(\nu) L_{-\nu}(\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)[L_{\nu_j}(\mu) + L_{-\nu_j}(\mu)] - \delta(\mu-\mu_1) + C(\mu), \quad \mu < 0,\ \mu_1 < 0, \tag{4.34}$$

whence

$$-\nu\Gamma_<(\nu) = \frac{1}{2\pi i}\int_0^1 d\nu_1\Gamma_<(\nu_1)\int_0^1 d\nu' \frac{X^-(\nu')}{Y^-(\nu';\nu')} I_{\nu'}(\nu) \times L_{-\nu_1}(-\nu') + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)\frac{1}{2\pi i}\int_0^1 d\nu' \times \frac{X^-(\nu')}{Y^-(\nu';\nu')} I_{\nu'}(\nu)[L_{\nu_j}(-\nu') + L_{-\nu_j}(-\nu')] - \frac{1}{2\pi i}\frac{X^-(-\mu_1)}{Y^-(-\mu_1;-\mu_1)} I_{-\mu_1}(\nu) + \frac{1}{2\pi i}\int_0^1 d\nu' \times \frac{X^-(\nu')}{Y^-(\nu';\nu')} I_{\nu'}(\nu)\, C(-\nu'), \quad \mu_1 < 0. \tag{4.35}$$

The additional conditions which determine $\Gamma_<^0(\nu_j)$ are

$$\int_0^1 d\nu\,\Gamma_<(\nu)\int_0^1 d\nu'\,\nu'^{l}\frac{X^-(\nu')}{Y^-(\nu';\nu')} L_{-\nu}(-\nu') + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_<^0(\nu_j)\int_0^1 d\nu'\,\nu'^{l}\frac{X^-(\nu')}{Y^-(\nu';\nu')}[L_{\nu_j}(-\nu') + L_{-\nu_j}(-\nu')] - (-\mu_1)^{l}\frac{X^-(-\mu_1)}{Y^-(-\mu_1;-\mu_1)} + \int_0^1 d\nu'\,\nu'^{l}\frac{X^-(\nu')}{Y^-(\nu';\nu')} C(-\nu') = 0,$$
$$l = 0, 1, \ldots, M. \tag{4.36}$$

5. MILNE PROBLEM FOR THE EXTERIOR OF A BLACK SPHERE

As a further illustration of the general formalism presented in Sec. 2 we consider the specific problem of determining the neutron angular density in the region exterior to a black sphere of radius r_s. By a black sphere we mean that the sphere is purely absorbing. Thus, the boundary condition for such a situation is

$$\Psi(r_s,\mu) = 0, \quad \mu > 0. \tag{5.1}$$

Further, we assume an isotropic source of the form

$$Q(r) = \lim_{R\to\infty}[q(R)/R^2]\delta(r - R), \quad R > r_s, \tag{5.2}$$

where $q(R)$ is to be chosen appropriately such that the spectral representation of the angular density Ψ_0 (due to the source) involves only one nonvanishing discrete normal mode. The choice of such a mode should become clear soon. We shall refer to this problem, together with the boundary condition (5.1) and the source function (5.2), as the exterior Milne problem.

Most of the steps involved in calculating $\Psi(r,\mu)$ for this problem are exactly parallel to the ones we encountered in the treatment of the interior problems. Therefore, to avoid repetition, we shall present only pertinent results, indicating only, wherever necessary, how those results were obtained. However, the source term for this problem engenders some extra steps which are discussed in detail in Appendix C.

Now in Eq. (2.1) let us put the explicit form of $Q(r)$, as given by Eq. (5.2), and rewrite it as

$$\Psi(r,\mu)\Theta(r - r_s) = \Psi_0(r,\mu) - \frac{1}{8\pi^2}\int_{-\infty}^{\infty} dK\, K^2 \int_{-1}^{1}\frac{dt}{1 + iKt} e^{iKrt\mu} \times J_0(Kr[(1-t^2)(1-\mu^2)]^{1/2}) \times [F(K,t) - C\sum_{l=0}^{N} b_l \frac{2l+1}{4\pi} P_l(t)\tilde{R}_l(K)], \tag{5.3}$$

where

$$\Psi_0(r,\mu) = \lim_{R\to\infty}\frac{1}{4\pi^2}\int_{-\infty}^{\infty} dK\, K^2\int_{-1}^{1}\frac{dt}{1 + iKt} e^{iKrt\mu} \times J_0(Kr[(1-t^2)(1-\mu^2)]^{1/2}) \times \left(\frac{2\pi q}{iKR} e^{iKR} + C\sum_{l=0}^{N} b_l \frac{2l+1}{4\pi} P_l(t)\, S_l^{(1)}(K)\right) \tag{5.4}$$

is the source angular density which has been separated out by virtue of the decomposition (2.12). Also we decomposed the amplitude $S_l(K)$ such that

$$S_l(K) = S_l^{(1)}(K) + S_l^{(2)}(K), \tag{5.5}$$

where $S_l^{(1)}(K)$ and $S_l^{(2)}(K)$ are determined by

$$\sum_{l=0}^{N} S_l^{(1),(2)}(K)\, A(l,l';K) = \pm\, 4\pi^2 q\, \frac{e^{\pm iKR}}{iKR} \int_{-1}^{1} dt\, \frac{P_l'(t)}{1+iKt}\,. \tag{5.6}$$

Here amplitudes $S_l^{(1)}(K)$ and $S_l^{(2)}(K)$ converge in the upper and the lower half of the complex K plane, respectively.

In Eq. (5.3) if we use the identity (2.16) to expand the product of the exponential and the Bessel functions in the second term on the right-hand side, express the whole integral in terms of the spherical Hankel functions of the first kind, consider the contour shown in Fig. 1, and imitate the procedure of Sec. 3, we get

$$\Psi(r,\mu)\Theta(r-r_s) = \Psi_0(r,\mu) + \frac{1}{2\pi i}\int_0^1 d\nu\, \Gamma_>(\nu) E_\nu^{(1)}(r,\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M} E_{\nu_j}^{(1)}(r,\mu)\,\Gamma_>^0(\nu_j), \tag{5.7}$$

where

$$E_\nu^{(1)}(r,\mu) = \frac{C}{2}\sum_{l=0}^{N} b_l(2l+1)g_l(\nu)\int_{-1}^{1} dt\, P_l(t) \times \sum_{n=0}^{\infty} i^n(2n+1)P_n(t)P_n(\mu)\, h_n^{(1)}\left(\frac{ir}{\nu}\right) \times\left(\mathcal{P}\frac{\nu}{\nu-t} + \pi i\nu\delta(\nu-t)\,\frac{A^+(l,0;\nu)+A^-(l,0;\nu)}{A^+(l,0;\nu)-A^-(l,0;\nu)}\right), \tag{5.8}$$

$$E_{\nu_j}^{(1)}(r,\mu) = \frac{C}{2}\sum_{l=0}^{N} b_l(2l+1)g_l(\nu_j)\int_{-1}^{1} dt\, \frac{\nu_j}{\nu_j - t}P_l(t) \times \sum_{n=0}^{\infty} i^n(2n+1)P_n(t)P_n(\mu)\, h_n^{(1)}\left(\frac{ir}{\nu_j}\right), \tag{5.9}$$

$$\Gamma_>(\nu) \equiv [\tilde{R}_0^+(\nu) - \tilde{R}_0^-(\nu)]/8\pi^2\nu^4, \tag{5.10}$$

and

$$\Gamma_>^0(\nu_j) = \lim_{\nu\to\nu_j}(\nu-\nu_j)\tilde{R}_0(\nu)/4\pi\nu_j^2. \tag{5.11}$$

An equation which determines $\Gamma_>(\nu)$ and $\Gamma_>^0(\nu_j)$ is

$$\Psi(r_s,\mu) = \Psi_0(r_s,\mu) + \frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_>(\nu)E_\nu^{(1)}(r_s,\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_>^0(\nu_j)E_{\nu_j}^{(1)}(r_s,\mu). \tag{5.12}$$

Again due to the regularity of the modes $E_\nu^{(1)}, E_{\nu_j}^{(1)}$, this equation is hard to manage analytically. However, an application of the reduction operator, defined by Eq. (2.38), to Eq. (5.3) will, in the same way as previously, permit us to produce a singular integral equation which can be solved. Before we consider that step, let us first present the spectral representation of the angular density Ψ_0 due to the source. Our standard procedure yields

$$\Psi_0(r,\mu) = \lim_{R\to\infty}\left(\frac{1}{2\pi i}\int_0^1 d\nu\, E_\nu(r,\mu)\,\frac{S_0^{(1)+}(\nu)-S_0^{(1)-}(\nu)}{8\pi^2\nu^4} + \frac{1}{2\pi i}\sum_{j=0}^{M} E_{\nu_j}(r,\mu)\lim_{z\to\nu_j}(z-\nu_j)\frac{S_0^{(1)}(z)}{4\pi^2\nu_j^2}\right), \tag{5.13}$$

where $E_\nu(r,\mu)$ and $E_{\nu_j}(r,\mu)$ are the normal modes defined by Eqs. (5.8) and (5.9), respectively. In the limit $R\to\infty$ this representation of Ψ_0 reduces to (see Appendix C for details)

$$\Psi_0(r,\mu) = \frac{1}{2\pi i}E_{\nu_0}(r,\mu)\eta(\nu_0), \tag{5.14}$$

where

$$\eta(\nu_0) = \ln\frac{\nu_0+1}{\nu_0-1} - \frac{1}{4\pi^2\nu_0^2}\sum_{l=0}^{N} f_l^{(0)}(\nu_0)A(l,0;\nu_0) \tag{5.15}$$

and the $f_l^{(0)}(\nu_0)$ are polynomials in ν_0 which are determined by the recurrence relation

$$\nu_0(1-cb_l)f_l^{(0)}(\nu_0) - \left(\frac{l+1}{2l+1}f_{l+1}^{(0)}(\nu_0) + \frac{l}{2l+1}f_{l-1}^{(0)}(\nu_0)\right) = 8\pi^2\nu_0^2\delta_{l0},\quad f_0^{(0)} \equiv 0. \tag{5.16}$$

Finally the spectral representation of the angular density, given by Eq. (5.7), now may be written in the form

$$\Psi(r,\mu) = \frac{1}{2\pi i}E_{\nu_0}(r,\mu)\eta(\nu_0) + \frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_>(\nu)E_\nu^{(1)}(r,\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_>^0(\nu_j)E_{\nu_j}^{(1)}(r,\mu). \tag{5.17}$$

6. APPLICATION OF THE REDUCTION OPERATOR TO THE EXTERIOR MILNE PROBLEM

In precisely the same way as we dealt with interior problems in Sec. 4, we apply the reduction operator $T_l^{(2)}$, defined by (2.38), to Eq. (5.3). If we parallel the procedure of that section here and use the results obtained in Appendix C for the external source, then we arrive at the following singular integral equation for expansion coefficients $\Gamma_>(\nu)$ and $\Gamma_>^0(\nu_j)$ in Eq. (5.17):

$$\frac{1}{2\pi i}\int_0^1 d\nu\,\Gamma_>(\nu)K_\nu(\mu) + \frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_>^0(\nu_j)K_{\nu_j}(\mu) + \frac{1}{2\pi i}[K_{\nu_0}(\mu)+K_{-\nu_0}(\mu)]\,\eta(\nu_0) = \Psi(r_s,\mu),\qquad \mu>0, \tag{6.1}$$

where

$$K_\nu(\mu) = \frac{1}{\pi r_s}\sum_{l=0}^{N} U_l^{(1)}\left(\frac{i}{\nu};r_s,\mu\right)g_l(\nu)\Big([B^+(l,0;\mu) - B^-(l,0;\mu)]\mathcal{P}\frac{\nu}{\nu+\mu} + \pi i[B^+(l,0;\mu)+B^-(l,0;\mu)] \times \nu\delta(\nu-\mu)\Big) \tag{6.2}$$

and

$$K_{\nu_0}(\mu) = \frac{1}{\pi r_s}\sum_{l=0}^{N} U_l^{(1)}\left(\frac{i}{\nu_0};r_s,\mu\right)g_l(\nu_0)[B^+(l,0;\mu) - B^-(l,0;\mu)]\frac{\nu_0}{\nu_0-\mu}\,. \tag{6.3}$$

The functions $U_l^{(1)}$ are, as before, defined by Eq. (4.8), while $B(l,0;z)$ are defined by Eq. (2.29) (with $K=i/z$).

The integral equation (5.18) for the boundary condition (5.1) has the solution

$$\nu\Gamma_>(\nu) = \frac{1}{2\pi i}\int_0^1 d\nu'\,\frac{\tilde{X}^-(\nu')}{\tilde{Y}^-(\nu';\nu')}\,\tilde{I}_{\nu'}(\nu)\left(\frac{1}{2\pi i}\sum_{j=0}^{M}\Gamma_>^0(\nu_j) \times K_{\nu_j}(\nu') + \frac{1}{2\pi i}[K_{\nu_0}(\nu')+K_{-\nu_0}(\nu')]\eta(\nu_0)\right),\qquad \nu>0, \tag{6.4}$$

where

$$\tilde{X}(z) = \frac{1}{(1-z)^{M+1}} \exp\left(\frac{1}{2\pi i}\int_0^1 \frac{d\nu}{\nu - z} \ln \frac{\tilde{Y}^+(\nu;\nu)}{\tilde{Y}^-(\nu;\nu)}\right), \quad (6.5)$$

$$\tilde{Y}(\nu;z) = \frac{1}{\pi r_s} \sum_{l=0}^{N} U_l^{(1)}\left(\frac{i}{\nu}; r_s, \mu\right) g_l(\nu) B(l,0;z), \quad (6.6)$$

and

$$\tilde{I}_{\nu'}(\nu) = \left(\frac{1}{\tilde{X}^-(\nu)} - \frac{1}{\tilde{X}^+(\nu)}\right) \mathcal{P} \frac{1}{\nu' - \nu} + \pi i \left(\frac{1}{\tilde{X}^-(\nu)} + \frac{1}{\tilde{X}^+(\nu)}\right) \times \delta(\nu - \nu'). \quad (6.7)$$

The additional conditions which determine $\Gamma^0_>(\nu_j)$ are

$$\sum_{j=0}^{M} \Gamma^0_>(\nu_j) \int_0^1 d\nu\, \nu^l \frac{\tilde{X}^-(\nu)}{\tilde{Y}^-(\nu;\nu)} K_{\nu_j}(\nu) = -\eta(\nu_0) \int_0^1 d\nu\, \nu^l \frac{\tilde{X}(\nu)}{\tilde{Y}(\nu;\nu)} [K_{\nu_0}(\nu) + K_{-\nu_0}(\nu)]. \quad (6.8)$$

We have precisely $M+1$ such equations from which all $M+1$ unknown discrete coefficients $\Gamma^0_>(\nu_j)$ may be determined. The continuum coefficients $\Gamma_>(\nu)$ are then exactly determined by Eq. (5.21).

ACKNOWLEDGMENTS

The author wishes to express his deep gratitude to Professor Kenneth M. Case for originally suggesting the problem and Professor William M. Irvine for many useful discussions. This work was supported in part by NASA Grant 22-010-023.

APPENDIX A: DERIVATION OF THE AUXILIARY LINEAR RELATIONS (2.31) AND (2.32)

Derivation of the auxiliary set of algebraic relations among $R_l^{(1)}$ and among $\tilde{R}_l$ [see Eqs. (2.32) and (2.31) in Sec. 2] follow exactly the same steps. We therefore present the derivation of Eq. (2.32) only.

Consider the set of linear equations (2.18)

$$\sum_{l=0}^{N} R_l^{(1)}(K) \left(\delta_{ll'} - Cb_l (l+\tfrac{1}{2}) \int_{-1}^{1} d\mu' \frac{P_l(\mu')P_{l'}(\mu')}{1+iK\mu'}\right) = -2\pi^2 r_s^2 \sum_{n=0}^{\infty} (-1)^n i^n (2n+1) h_n^{(1)}(Kr_s) \int_{-1}^{1} dt \times \frac{P_{l'}(t)P_n(t)}{1+iKt} \int_{-1}^{1} dt'\, t'\, P_n(t')\Psi(r_s,t'). \quad (A1)$$

In order to deduce Eq. (2.32) from Eq. (A1) the basic idea is to use the completeness and orthogonality properties of the Legendre polynomia, i.e.,

$$\sum_{l=0}^{\infty} (l+\tfrac{1}{2}) P_l(\mu) P_l(\mu') = \delta(\mu - \mu') \quad (A2)$$

and

$$\int_{-1}^{1} d\mu\, P_l(\mu) P_{l'}(\mu) = \frac{2}{2l+1}\delta_{ll'}. \quad (A3)$$

Thus, if we multiply Eq. (A1) by $(l'+\frac{1}{2})P_{l'}(\mu)$ and sum over l', then using the completeness relation (A2), we obtain

$$\sum_{l=0}^{N} R_l^{(1)}(K)(l+\tfrac{1}{2})(1+iK\mu - Cb_l)P_l(\mu) = -2\pi r_s^2 \sum_{n=0}^{\infty} (-1)^n i^n (2n+1) h_n^{(1)}(Kr_s) P_n(\mu) \times \int_{-1}^{1} dt\, t P_n(t) \Psi(r_s,t). \quad (A4)$$

An application of the orthogonality property (A3) to Eq. (A4) yields

$$\sum_{l=0}^{N} R_l^{(1)}(K) \left[(1-Cb_l)\delta_{nl} + iK\left(\frac{l+1}{2n+1}\delta_{n\,l+1} + \frac{l}{2n+1}\delta_{n\,l-1}\right)\right] = -4\pi^2 r_s^2 (-1)^n i^n h_n^{(1)}(Kr_s) \times \int_{-1}^{1} dt\, t P_n(t)\Psi(r_s,t). \quad (A5)$$

Now, dividing Eq. (A5) by $(-1)^n i^n h_n^{(1)}(Kr_s)(1-iK\mu)[(n+\frac{1}{2})P_n(\mu)]^{-1}$ summing over n from 0 to ∞ and applying the completeness relation (A2), we get

$$\sum_{l=0}^{N} \frac{R_l^{(1)}(K)}{1-iK\mu} \left[i^l (1-Cb_l)(2l+1) \frac{P_l(\mu)}{h_l^{(1)}(Kr_s)} + iK\left((l+1)i^{l+1} \frac{P_{l+1}(\mu)}{h_{l+1}^{(1)}(Kr_s)} + l\, i^{l-1} \frac{P_{l-1}(\mu)}{h_{l-1}^{(1)}(Kr_s)}\right)\right] = -8\pi^2 r_s^2 \mu \frac{\Psi(r_s,\mu)}{1-iK\mu}. \quad (A6)$$

The result of integration with respect to μ of this equation is precisely the set of linear equations (2.32).

APPENDIX B: EXPLICIT EXPRESSIONS FOR CONTINUUM AND DISCRETE NORMAL AMPLITUDES DUE TO THE INTERIOR SOURCE

Consider the factorization equation (2.24). For the source amplitudes $S_l(K)$ we have

$$S_l(z) = g_l(z) S_0(z) + f_l(z), \quad (B1)$$

where the $g_l(z)$ are determined by the recurrence relation (2.25), while the $f_l(z)$ for a point source are determined by

$$z(1-Cb_l)f_l(z) - \left(\frac{l+1}{2l+1} f_{l+1}(z) + \frac{l}{2l+1} f_{l-1}(z)\right) = 4\pi z q \delta_{l0}, \quad f_0 \equiv 0 \quad (B2)$$

In the linear relations (2.14) among $S_l(z)$ if we apply the factorization formula (B1) for the amplitudes $S_l(z)$, we obtain (for $l'=0$)

$$\Lambda(z) S_0(z) = 2\pi q z \int_{-1}^{1} \frac{dt}{z-t} - \sum_{l=0}^{N} f_l(z) A(l,0;z), \quad (B3)$$

where

$$\Lambda(z) = \sum_{l=0}^{N} g_l(z) A(l,0;z). \quad (B4)$$

Then, across the cut $-1 \le \nu \le 1$,

$$S_0^+(\nu) - S_0^-(\nu) = -2\pi q \int_{-1}^{1} dt \left[\left(\frac{1}{\Lambda^+(\nu)} - \frac{1}{\Lambda^-(\nu)}\right) \mathcal{P} \frac{\nu}{\nu - t} + \pi i \left(\frac{1}{\Lambda^+(\nu)} + \frac{1}{\Lambda^-(\nu)}\right) \nu \delta(\nu - t)\right] - \sum_{l=0}^{N} f_l(\nu) \left(\frac{A^+(l,0;\nu)}{\Lambda^+(\nu)} - \frac{A^-(l,0;\nu)}{\Lambda^-(\nu)}\right) \quad (B5)$$

and at simple zeros ν_j of $\Lambda(z)$, i.e.,

$$\Lambda(\nu_j) = 0, \quad j = 0, 1, \ldots, M, \quad (B6)$$

$$\lim_{z\to\nu_j}(z-\nu_j)S_0(z)=\frac{1}{\Lambda'(\nu_j)}\left(2\pi q\int_{-1}^{1}dt\,\frac{\nu_j}{\nu_j-t}-\sum_{l=0}^{N}f_l(\nu_j)A(l,0;\nu_j)\right),\quad \text{(B7)}$$

where the prime on $\Lambda'(\nu_j)$ denotes the derivative of $\Lambda(z)$ evaluated at $z=\nu_j$.

APPENDIX C: ANGULAR DENSITY $\Psi_0(r,\mu)$ DUE TO THE SOURCE IN THE EXTERIOR MILNE PROBLEM

To see how the representation of $\Psi_0(r,\mu)$ given by Eq. (5.13) reduces in the limit $R\to\infty$ to Eq. (5.14), which involves only one discrete normal mode $E_{\nu_0}(r,\mu)$, consider the set of linear relations (5.6) among $S_l^{(1)}(K)$. In the $z=i/K$ plane

$$\sum_{l=0}^{N}S_l^{(1)}(z)A(l,l';z)=-4\pi^2q\,e^{-R/z}\,\frac{z^2}{R}\int_{-1}^{1}dt\,\frac{P_{l'}(t)}{z-t}.\quad \text{(C1)}$$

Since, by virtue of the factorization formula (2.24),

$$S_l^{(1)}(z)=g_l(z)\,S_0^{(1)}(z)+f_l^{(1)}(z),\quad \text{(C2)}$$

where $f_l^{(1)}(z)$ are functions which are determined by the recurrence relation

$$z(1-cb_l)f_l^{(1)}(z)-\left[\frac{l+1}{2l+1}f_{l+1}^{(1)}(z)+\frac{l}{2l+1}f_{l-1}^{(1)}(z)\right]=-8\pi^2z^2q\,\frac{e^{-R/z}}{R}\,\delta_{l0},\quad f_0^{(1)}\equiv 0,\quad \text{(C3)}$$

we may, using Eq. (C2), re-express Eq. (C1) (for $l'=0$) as

$$\Lambda(z)S_0^{(1)}(z)=-\left(4\pi^2q\,e^{-R/z}\,\frac{z}{R}\int_{-1}^{1}dt\,\frac{z}{z-t}+\sum_{l=0}^{N}f_l^{(1)}(z)\times A(l,0;z)\right).\quad \text{(C4)}$$

In particular at the zeros of $\Lambda(z)$, i.e.,

$$\Lambda(\nu_j)=\sum_{l=0}^{N}g_l(\nu_j)A(l,0;\nu_j)=0,\quad j=0,1,\ldots,M,\quad \text{(C5)}$$

the discrete normal amplitudes are given by

$$\lim_{z\to\nu_j}(z-\nu_j)S_0^{(1)}(z)=-\frac{1}{\Lambda'(\nu_j)}\left[4\pi^2q\,e^{-R/\nu_j}\,\frac{\nu_j}{R}\int_{-1}^{1}dt\,\frac{\nu_j}{\nu_j-t}+\sum_{l=0}^{N}f_l^{(1)}(\nu_j)A(l,0;\nu_j)\right],\quad j=0,1,\ldots,M.\quad \text{(C6)}$$

Now, if we partially order ν_j, s such that

$$\mathrm{Re}\,\nu_0/|\nu_0|>\mathrm{Re}\,\nu_1/|\nu_1|>\cdots>\mathrm{Re}\,\nu_M/|\nu_M|,\quad \text{(C7)}$$

where Re means the real part, and choose q so that

$$q=-\,\mathrm{Re}^{-R/\nu_0}\,\Lambda'(\nu_0),\quad \text{(C8)}$$

then Eq. (C6) in the limit $R\to\infty$ reduces to

$$\lim_{z\to\infty}\left(\lim_{z\to\nu_j}(z-\nu_j)S_0^{(1)}(z)\right)=\delta_{j0}\,4\pi^2\nu_0^2\int_{-1}^{1}\frac{dt}{\nu_0-t}-\frac{1}{\Lambda'(\nu_j)}\lim_{R\to\infty}\sum_{l=0}^{N}f_l^{(1)}(\nu_j)\times A(l,0,\nu_j).\quad \text{(C9)}$$

The second term on the right-hand side of this equation is further simplified by noting that in the limit $R\to\infty$ the recurrence relation (C3), for $z=\nu_j$ and q as given by Eq. (C8), becomes

$$\lim_{R\to\infty}\left[\nu_j(1-cb_l)f_l^{(1)}(\nu_j)-\left(\frac{l+1}{2l+1}f_{l+1}^{(1)}(\nu_j)+\frac{l}{2l+1}f_{l-1}^{(1)}(\nu_j)\right)\right]=8\pi^2\nu_j^2\,\delta_{l0}\delta_{j0}\Lambda'(\nu_j);\ \lim_{R\to\infty}f_0^{(1)}\equiv 0.\quad \text{(C10)}$$

From this we conclude that

$$\lim_{R\to\infty}f_l^{(1)}(\nu_j)=\delta_{j0}f_l^{(0)}(\nu_0)\Lambda'(\nu_0),\quad \text{(C11)}$$

where $f_l^{(0)}(\nu_0)$ are polynomials in ν_0 determined by the recurrence relation

$$\nu_0(1-cb_l)f_l^{(0)}(\nu_0)-\left(\frac{l+1}{2l+1}f_{l+1}^{(0)}(\nu_0)+\frac{l}{2l+1}f_{l-1}^{(0)}(\nu_0)\right)=8\pi^2\nu_0^2\delta_{l0},\quad f_0^{(0)}\equiv 0.\quad \text{(C12)}$$

Using the relation (C11) in Eq. (C9) for the discrete normal amplitudes, we get

$$\lim_{R\to\infty}\left(\lim_{z\to\nu_j}(z-\nu_j)S_0^{(1)}(z)\right)=\delta_{j0}\left(4\pi^2\nu_0^2\ln\frac{\nu_0+1}{\nu_0-1}-\sum_{l=0}^{N}f_l^{(0)}(\nu_0)A(l,o;\nu_0)\right).\quad \text{(C13)}$$

Let us also note that with the choice of q, as given by Eq. (C8), all the continuum normal amplitudes in the spectral representation of $\Psi_0(r,\mu)$ [see Eq. (5.13)] vanish, i.e.,

$$\lim_{R\to\infty}[S_0^+(\nu)-S_0^-(\nu)]=0.\quad \text{(C14)}$$

---

* Contribution No. 109 of the Five College Observatories.

1 K. M. Case, R. Zelazny, and M. Kanal, J. Math Phys. **11**, 223 (1970).
2 K. M. Case, "On the Boundary-Value Problems of Linear Transport Theory," report (Department of Physics, University of Michigan, 1967).
3 K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
4 M. Kanal, J. Math. Phys. **11**, 3042 (1970).
5 B. Davison, *Neutron Transport Theory* (Oxford U.P., London, 1957).
6 G. J. Mitsis, Argonne National Laboratory Report ANL-6787 (1963).
7 R. C. Erdmann and C. E. Siewert, J. Math. Phys. **9**, 81 (1968).
8 G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1922).
9 E. Inönü, J. Math. Phys. 11, 568 (1970).
10 E. D. Rainville, *Special Functions* (Macmillan, New York, 1960).
11 N. I. Muskhelishvili, *Singular Integral Equations* (Noordoff, Groningen, 1946).

Analytic Renormalization of the Exponential Interaction: The Three-Point Time-Ordered Product with Minimum Light Cone Singularity

M. Daniel

Department of Theoretical Physics, University of Oxford, Oxford, England

and

P. K. Mitter*

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(29 November 1971)

A method of analytic renormalization is developed to define the three-point time-ordered product of massless fields of exponential type as a strictly localizable distribution. The uniqueness property, known for the two-point T-product, is verified for the three-point product, for a special choice of fine renormalization. It is characterized by minimum singularity on the light cone: There are no delta function type singularities concentrated on the surface $x_1 = x_2 = x_3$.

1. INTRODUCTION

The difficulties of "nonrenormalizable" field theories, at least at the level of perturbation theory, are well known although, at our present level of knowledge, no fundamental objection to (at least a class of) such theories is known.[1] At the perturbative level, such theories (e.g., $:\phi^n: n \geqq 5$) can indeed be subjected to "axiomatic renormalization",[2,3] i.e., a Lorentz invariant extension procedure of time ordered products consistent with axiomatic structure (unitarity, causality);[4] however, such procedures are not physically compelling for these theories due to the greater lack of uniqueness of extensions than is the case in renormalizable theories.

In this context, Lehmann and Pohlmeyer[5] made an interesting observation in connection with the two-point time-ordered product of the interaction Lagrangian:

$$\mathcal{L}_1(x) = :[\exp[g\phi(x)] - 1]:, \tag{1.1}$$

where $\phi(x)$ is a free scalar field on four-dimensional space–time. They observed that although $\mathcal{L}_1(x)$ leads to a conventionally nonrenormalizable theory, each order of perturbation in $\mathcal{L}_1(x)$ furnishing an infinite set of divergent Feynman graphs, there existed a unique definition of $T[\mathcal{L}_1(x_1)\mathcal{L}_1(x_2)]$ as a distribution in the sense of Jaffe.[1] The uniqueness was characterized by minimum singularity on the light cone corresponding to a certain regularity on the surface $x_1 = x_2$. The question naturally arises as to whether higher-point time-ordered products can be defined with minimum light cone singularity. Since the only ambiguity in the construction of time-ordered products, consistent with axiomatic structure, is in the form of distributions concentrated on surfaces of coinciding points,[6] a construction with minimum singularity, satisfying a regularity condition[7] on such surfaces, would be automatically unique and would not entail the addition of arbitrary finite counter terms to (1.1). If the question has an affirmative answer, then one might use the regularity condition at the light cone vertex as a boundary condition to fix the dynamics of at least a privileged class of nonrenormalizable fields. Furthermore, the analysis of (1.1) from this viewpoint may turn out to be of physical relevance, since nonrenormalizable theories of physical interest can be shown to contain such structures.[8]

In this paper we report on some progress in this direction. In an earlier paper[8] we showed, by a method different from that of Ref. 5, and with the restriction that $\phi(x)$ was massless, that the two-point time-ordered product $T(\mathcal{L}_1(x_1)\mathcal{L}_1(x_2))$ could be defined uniquely as a distribution with minimum light cone singularity. We presented a method of analytic renormalization, motivated by (but different from) that of Speer,[9,10] whereby the divergences of an infinite set of Feynman graphs were removed and a minimal extension was secured. In the present paper, our renormalization scheme is extended [with $\phi(x)$ a massless field] to the case of the three-point time-ordered product $T(\mathcal{L}(x_1)\mathcal{L}(x_2)\mathcal{L}(x_3))$ which is again a formally infinite set of divergent Feynman graphs. We prove that there exists a choice of finite renormalization, such that the T product enjoys minimum light cone singularity. For simplicity we have considered the case (1.1), although our results go through for more general fields of exponential type.

Before we delve into the details, it is worth noting that the regularity condition on surfaces $x_1 = x_2 = \cdots = x_n$ is nontrivial. This is because, apart from the case of a super-renormalizable field theory for which the condition is satisfied for all but a finite number of Feynman graphs,[11] no finite sum of renormalized Feynman graphs can enjoy this property due essentially to the presence of logarithms in renormalized perturbation theory. For the case at hand, we are exploiting a loophole in that we have a suitably convergent *infinite* series of (albeit divergent) Feynman graphs. The extension with minimum singularity then leads to a precise definition of minimal renormalization.

The material of this paper is distributed as follows: In Sec. 2 we review, for the convenience of the reader, our method of analytic renormalization[8] applied to the case of the two-point T product. Section 3 is devoted to the replacement of the general term, in the formal Wick expansion of $T(\mathcal{L}(x_1)\mathcal{L}(x_2)\mathcal{L}(x_3))$ in an infinite series of (divergent) Feynman graphs, by an analytically regulated object, a massless generalized Feynman amplitude (GFA) in the sense of Speer[9,10,12] which (i) exists as a well-defined tempered distribution and for which (ii) individual divergences are separately tracked with complex parameters. (The method of regulating with complex parameters is specially helpful for a massless theory.[12] Once this is done, we can define the regulated three-point T product through a series of GFA's converging to a strictly localizable distribution.[1] The next step, taken in Sec. 4, is to generalize the renormalization procedure of Sec. 2 to the three-point T product. The method is to give an analytic decomposition of the regulated object into a sum of regular and singular parts. Analytic continuation of the regular part to the (physical) value of the complex parameters gives a renormalization. The consistency of our procedure is established, through a verification of the unitarity–causality properties of T products. In Sec. 5 it is proved, for the three-point T product, that a finite renormalization may be implemented, so as to secure, precisely as in the case of the two-point T product, a unique extension characterized by a decrease property (in momentum space) for the real part corresponding to minimum singularity on the (x space) light cone. We show the absence of delta function type singularities concentrated on the surface $x_1 = x_2 = x_3$.

An entirely different approach to this problem has been taken recently by Pohlmeyer.[13] We wish to emphasize that our principal motivation in undertaking this analysis has been to extend methods originating in renormalization theory to Lagrangians of exponential type.

2. THE TWO-POINT FUNCTION AS A DISTRIBUTION

We begin by reviewing how the two-point T product $T(\mathcal{L}_1(x_1)\mathcal{L}_1(x_2))$ may be defined as an (operator valued) generalized function in a space of localizable generalized functions. Throughout we shall consider the case: when $\phi(x)$, (1.1) is a massless scalar field.

Through a formal application of Wick's theorem, we get

$$T(\mathcal{L}_1(x_1)\mathcal{L}_1(x_2)) = \sum_{i_1,i_2=0}^{\infty} \mathcal{T}_2(x_1,x_2) g^{i_1+i_2} \frac{:\phi(x_1)^{i_1}\phi(x_2)^{i_2}:}{i_1!\,i_2!} \qquad (2.1)$$

and for connected contributions,

$$\mathcal{T}_2(x_1x_2) = \sum_{\nu=1}^{\infty} \frac{(g^2)^\nu}{\nu!}\{(4\pi^2)[-(x_1-x_2)^2 + i0]\}^{-\nu} \quad (2.2)$$

and the metric is $g_{\mu\nu} = (1,-1,-1,-1)$. Our task is to define the formal expression (2.2) as a functional in $\mathcal{C}'_g$, a (Jaffe) space of strictly localizable generalized functions. We shall follow the method given in Ref. 8.

In lieu of (2.2), we consider

$$\mathcal{T}_2(\lambda)(x) = \sum_{\nu=1}^{\infty} \frac{(g^2)^{(1-\lambda)\nu}}{\nu!\,(4\pi^2)^\nu} (-x^2 + i0)^{(\lambda-1)\nu} \quad (2.3)$$

and the complex parameter[14] λ is restricted to the region

$$\Lambda = \{\lambda \mid 0 < \mathrm{Re}\,\lambda < 1, 0 < \mathrm{Im}\,\lambda < \alpha\}. \qquad (2.4)$$

It is readily seen that (2.3), as a series of meromorphic distributions in $\mathcal{S}'(R^4)$, converges in the topology of the space $\mathcal{C}'_g(R^4)$ chosen with indicator function $g(t)$ of order of growth ρ, $\frac{1}{3} < \rho < \frac{1}{2}$, which is consistent with strict localizability.[1] We obtain from (2.3), by continuity of the Fourier operator,

$$\tilde{\mathcal{T}}_2(\lambda)(p) = -i \sum_{\nu=1}^{\infty} \frac{(g^2)^{(1-\lambda)\nu}}{\nu![(4\pi)^2]^\nu} \times \frac{\Gamma((\lambda-1)\nu+2)}{\Gamma((1-\lambda)\nu)} (-p^2 - io)^{(1-\lambda)\nu-2} \quad (2.5)$$

as an element of $\mathfrak{M}'_g(R^4)$.[1] Then $\lambda = 0$ is a singular point for all terms (excepting the first) of the series (2.5). The νth term has poles at $\lambda = 1 - (k+2)/\nu$ ($k = 0, 1, 2, \ldots$), i.e., we have a pole at $\lambda = 0$ plus a sequence of neighboring poles. As ν increases, the neighboring poles approach $\lambda = 0$, so that $\tilde{\mathcal{T}}_2(\lambda)(p)$ has a *nonisolated* singularity at $\lambda = 0$.

The λ singularity in (2.5) can be extracted by casting (2.5) into an integral representation and analyzing its singularity[8] at $\lambda = 0$. To this end we introduce an analytic interpolation for the general term in (2.5) with the replacement $\nu \to z$.

Now (2.5) has the following integral representation:

$$\tilde{\mathcal{T}}_2(\lambda)(p) = \tfrac{1}{2}\int_\Gamma dz\, \frac{\cot\pi z\,(g^2)^{(1-\lambda)z}2^{2\lambda z}}{[(4\pi)^2]^z\,\Gamma(z+1)} \times \frac{\Gamma(2-(1-\lambda)z)}{\Gamma((1-\lambda)z)} (-p^2 - i0)^{(1-\lambda)z-2}, \quad (2.6)$$

with the contour Γ shown in Fig. 1 encircling the positive real axis, counterclockwise.

The singularity of (2.6) at $\lambda = 0$ arises from the pinch of the contour Γ between the moving z-plane poles (of the integrand) $Z^{(n)}(\lambda) = n/(1-\lambda)$, $n = 2, 3, 4, \cdots$, and the fixed poles (arising from $\cot\pi z$) at $n = 2, 3, \cdots$. This is illustrated in Fig. 2.

We then deform $\Gamma \to \bar{\Gamma}$ so as to enclose the $z^{(n)}(\lambda)$ pole and compute the discontinuity (Fig. 3), thus obtaining the following analytic decomposition with $\lambda \in \Lambda$:

$$\tilde{\mathcal{T}}_2(\lambda)(p) = \tilde{\mathcal{T}}_2(\lambda)(p)_R + \Delta\tilde{\mathcal{T}}_2(\lambda)(p), \qquad (2.7)$$

where $\tilde{\mathcal{T}}_2(\lambda)_R$ is obtained from (2.6) with the replacement $\Gamma \to \bar{\Gamma}$ and

$$\Delta\tilde{\mathcal{T}}_2(\lambda)(p) = \pi i \sum_{n=2}^{\infty} \times \frac{(-1)^n \cot[\pi n/(1-\lambda)](g^2)^n[2^{2\lambda}/(4\pi)^2]^{n/1-\lambda}}{\Gamma(n-1)\Gamma(n)\Gamma[(n/(1-\lambda)]+1)} \times (-p^2)^{n-2}. \qquad (2.8)$$

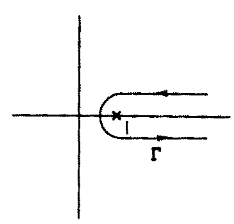

FIG. 1. The contour Γ.

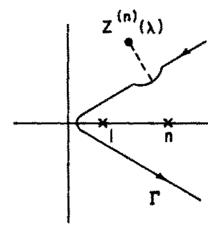

FIG. 2. Pinching of the contour Γ.

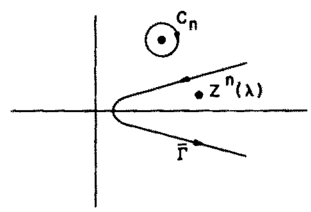

FIG. 3. Contour splitting.

We observe that $\tilde{T}_2(\lambda)_R$ is analytic in the neighborhood of $\lambda = 0$. As we continue to $\lambda = 0$, the moving poles $z^{(n)}(\lambda)$ merely coalesce with the fixed poles at $z = n$, and no further contour deformation is necessary, hence, analyticity.

The piece $\Delta\tilde{T}_2(\lambda)$, which has a nonisolated singularity at $\lambda = 0$, is readily seen to be an entire function of p^2 of order $\frac{1}{3}$. Hence its Fourier transform has support concentrated at $x = 0$. Thus $T_2(\lambda)$ is well defined at $\lambda = 0$ as a continuous linear functional on $\mathcal{C}_0(R^4)$, the subspace of test functions which vanish, together with all derivatives, at $x = 0$. It may now be defined at $\lambda = 0$ as an element of $\mathcal{C}'_g(R^4)$ through the following extension:

$$T_2(\mathbf{x}) = \mathfrak{F}^{-1}[\tilde{T}_2(\mathbf{p})], \tag{2.9}$$

where

$$\tilde{T}_2(\mathbf{p}) = \operatorname*{anal.cont.}_{\lambda\to 0}\, T_2(\lambda)(\mathbf{p})_R \tag{2.10}$$

$$= \tfrac{1}{2}\int_{\bar{\Gamma}} dz\, \frac{\cot\pi z}{[(4\pi)^2]^z}\, \frac{(g^2)^z}{\Gamma(z+1)}\, \frac{\Gamma(2-z)}{\Gamma(z)}\,(-p^2 - i\,0)^{z-2} \tag{2.11}$$

is a Lorentz invariant generalized function in $\mathfrak{M}'_g(R^4)$. In principle, we are free to add to our choice $T_2(\mathbf{x})$, any functional $f(\Box)\,\delta(x)$ concentrated at $x = 0$ [$f(s)$ an entire function of order less than one half]. However, the choice (2.9)–(2.11) constitutes a "*minimal*" extension. To see this we consider, for $p^2 > 0$,

$$\mathrm{Re}[i\tilde{T}_2(p)] = i\int_{\bar{\Gamma}} dz\, \frac{(g^2)^z[(4\pi)^2]^{-z}}{\sin\pi z\,\Gamma(z+1)}\, \frac{\Gamma(2-z)}{\Gamma(z)}\,(p^2)^{z-2}. \tag{2.12}$$

The contour $\bar{\Gamma}$ in (2.12) can be opened up parallel to the imaginary axis without losing convergence. Then (2.12) is rapidly decreasing as $p^2 \to +\infty$. In fact the integral can be recognized as a Meijer function G^{20}_{03} with $(g/4\pi)^2p^2$ as argument and which decreases exponentially [see Ref. 15, p. 312] as $p^2 \to +\infty$. This implies, by a space smearing[5] argument that $\mathrm{Re}[iT_2(x)]$ is free of δ-function type singularities concentrated at $x = 0$. Since $\mathrm{Im}[iT_2(x)]$ can be uniquely defined in $\mathcal{C}'_g(R^4)$, the extension (2.9)–(2.11) is unique by minimum light cone singularity, the criterion of Ref. 5.

In the subsequent sections we will show how these results may be generalized to the case of the three-point function.

3. MASSLESS GENERALIZED FEYNMAN AMPLITUDES

(A) We start with the *formal* expression, obtained through Wick's theorem:

$$T(\mathcal{L}_1(x_1)\mathcal{L}_1(x_2)\mathcal{L}_1(x_3)) = \sum_{i_1,i_2,i_3=0}^{\infty} T_3(x_1x_2x_3)\, g^{\sum_1^3 i_l} \times \frac{:\phi(x_1)^{i_1}\phi(x_2)^{i_2}\phi(x_3)^{i_3}:}{i_1!\,i_2!\,i_3!}. \tag{3.1}$$

By restricting ourselves to connected contributions, any coefficient function in (3.1) may be expressed as

$$T_3(x_1x_2x_3) = \sum_{\nu_1\nu_2\nu_3=1}^{\infty} \frac{(g^2)^{\sum_1^3\nu_l}}{\Pi_{l=1}^3[\nu_l!]}\, T_3^{(\nu_1\cdots\nu_3)}(x_1x_2x_3), \tag{3.2}$$

where

$$T_3^{(\nu_1\nu_2\nu_3)}(x_1x_2x_3) = \prod_{l=1}^{3}\{(4\pi^2)[-(x_{l_i} - x_{l_f})^2 + i0]\}^{-\nu_l}. \tag{3.3}$$

Each term of (3.2) may be realized as a connected graph $G(\{V\}, \mathcal{L})$ with a set of vertices $\{V\}$ with $\#\{V\} = 3$ and set of lines $\mathcal{L}$, with $L = \#\{\mathcal{L}\} = 3$. If we make a correspondence with Feynman graphs each line $l \in \mathcal{L}$ is a *multiplet* of ν_l massless lines.[12] The formal nature of (3.1)–(3.3) arises from the lack of definition of (3.3), which involves products of distributions.

$T_3^{(\nu_1\cdots\nu_3)}$ is a Feynman amplitude with massless lines. Thus our first step will be to replace it by an analytically regulated object, a generalized Feynman amplitude (GFA) in the sense of Speer[9,10,12] which (i) exists as a distribution, and (ii) reproduces (3.3) formally when regulating parameters are assigned certain values (Sec. 3B). We will then analytically continue the massless GFA to a region (Sec. 3C) in which (3.2), with $T_3^{(\nu_1\cdots\nu_3)}$ replaced by GFA, represents a series of distribution in $\mathcal{S}'(R^{4\times3})$ converging to an element of $\mathcal{C}'_g(R^{4\times3})$ (Sec. 3D). At this stage we will have achieved a definition of the regulated version of (3.2) in a form suitable for renormalization.

(B) In accordance with our program we will introduce in this subsection an analytically regulated version of the massless Feynman amplitude (3.3), avoiding difficulties with infrared divergences.[16] In this subsection $\nu_1\cdots\nu_3$ in (3.3) are arbitrary positive integers but fixed.

In lieu of the undefined factors in (3.3) we introduce

$$\Delta^{(\nu_l)}(\lambda_l) = \lim_{\eta\downarrow 0}\,[1/(4\pi^2)^{\nu_l}] \times[-(x_{l_i} - x_{l_f})^2 + i\eta\|x_{l_i} - x_{l_f}\|^2]^{(\lambda_l-1)\nu_l}, \tag{3.4}$$

which is a meromorphic distribution[17] in $\mathcal{S}'(R^{4\times2})$ (λ_l: complex parameters) and formally reproduces the undefined factors in (3.3) for $\lambda_l = 0$. $\|\ \|$ is the Euclidean norm in R^4. We then introduce a further real (>0) parameter r,

$$\Delta^{(\nu_l)}(\lambda_l)(x) = \lim_{\eta\downarrow 0}\,\lim_{r\downarrow 0}\Delta^{(\nu_l)}_{\eta,r}(\lambda_l)(x) \tag{3.5}$$

with the rhs of (3.5) defined as the inverse F.T. (in R^4) of

$$\tilde{\Delta}^{(\nu_l)}_{\eta,r}(\lambda_l)(p) = \frac{i(1-i\eta)(1+\eta^2)^{1/2}e^{[(\lambda_l-1)\nu_l](\pi i/2)}}{(4\pi^2)^{\nu_l}}\cdot\frac{2^{[(\lambda_l-1)\nu_l+1]}}{\Gamma((1-\lambda_l)\nu_l)} \times\int_r^{\infty} d\alpha_l\,\alpha_l^{[(\lambda_l-1)\nu_l]+1}\exp[i\alpha_l(p^2 + i\eta\|p\|^2)] \tag{3.6}$$

with $\mathrm{Re}\lambda_l < 1$.

$\Delta^{(\nu_l)}_{\eta,r}(\lambda_l)(x_{l_i} - x_{l_f})$ is continuous and bounded, the product

$$T^{\boldsymbol{\nu}}_{\eta,r}(\boldsymbol{\lambda})(\mathbf{x}) = \prod_{l=1}^{3}\Delta^{(\nu_l)}_{\eta,r}(\lambda_l)(x_{l_i} - x_{l_f}) \tag{3.7}$$

is well defined and may be evaluated by standard methods[9] to get the regulated Feynman amplitude in α space:

$$\tilde{T}^{\boldsymbol{\nu}}_{\eta,r}(\boldsymbol{\lambda})(\mathbf{p}) = (8\pi)^2\,\delta\left(\sum_{i=1}^{3} p_i\right) \times\prod_{l=1}^{3}\left(\frac{2^{2[(\lambda_l-1)\nu_l}e^{(\pi i/2[(\lambda_l-1)\nu_l]}}{(4\pi^2)^{\nu_l}\Gamma((1-\lambda_l)\nu_l)}\right)\cdot\frac{1}{(1+\eta^2)^{1/2}(1+i\eta)}$$

$$\times \int_r^\infty d\alpha_1 \int_r^\infty d\alpha_2 \int_r^\infty d\alpha_3 \frac{\Pi_{l=1}^3 (\alpha_l^{(\lambda_l-1)\nu_l+1})}{(\alpha_1+\alpha_2+\alpha_3)^2}$$

$$\times \exp\left[\frac{i}{1+\eta^2}\left(\frac{1}{(\alpha_1+\alpha_2+\alpha_3)}\right.\right.$$

$$\times [(\alpha_3\alpha_1 p_1^2 + \alpha_1\alpha_2 p_2^2 + \alpha_2\alpha_3 p_3^2)$$

$$\left.\left. + i\eta(\alpha_3\alpha_1\|p_1\|^2 + \alpha_1\alpha_2\|p_2\|^2 + \alpha_2\alpha_3\|p_3\|^2)]\right)\right]. \tag{3.8}$$

In order to take the $r \to 0+$ limit on (3.8) in a restricted region of the regulating parameters $\boldsymbol{\lambda}$, we introduce[9,12] a sectorial decomposition of α space. We write (3.8) in the form

$$\tilde{T}^{\nu}_{\eta,r}(\boldsymbol{\lambda}) = \sum_\rho \tilde{T}^{\nu}_{\eta,r}(\boldsymbol{\lambda})_\rho, \tag{3.9}$$

where ρ is a one-to-one map of $\{1,2,3\}$ onto $\mathcal{L}$ and π_ρ is the sector in the α space given by

$$\pi_\rho = \{\boldsymbol{\alpha} \mid 0 \le \alpha_{\rho(1)} \le \alpha_{\rho(2)} \le \alpha_{\rho(3)} \le \infty\}, \tag{3.10}$$

and introduce sector coordinates $\alpha_{\rho(l)} = t_3 \cdots t_l$ such that $0 \le t_3 \le \infty$, $0 \le t_l \le 1$ $(l = 1, 2)$ with Jacobian $\Pi_1^3 t_l^{l-1}$. Then the "α integral" in (3.8) in the sector π_ρ is

$$(\alpha\text{ integral})_\rho = \int_r^\infty dt_3\, t_3^{\mu_3^\rho+3} \int^1 dt_2 \int^1 dt_1 \prod_{l=1}^{2} \{t_l^{\mu_l^\rho+2l-1}\}$$

$$\times [E(t_1,t_2)]^{-2} \times \exp\left(i\frac{t_3}{1+\eta^2}\cdot\frac{t_2}{E(t_1t_2)}\right.$$

$$\left.\times [P_1(\mathbf{t};\mathbf{p}) + i\eta P_2(\mathbf{t};\mathbf{p})]\right), \tag{3.11}$$

where

(i) $\mu_l^\rho \equiv \sum_{l'=1} [(\lambda_{\rho(l')} - 1)\nu_{\rho(l')}]$,

(ii) $E(t_1,t_2) \equiv 1 + t_2 + t_1 t_2$, (3.12)

(iii) $P_1^\rho \equiv p_{\rho(3)}^2 + t_1t_2 p_{\rho(2)}^2 + t_1 p_{\rho(1)}^2$,

$P_2 \equiv \|p_{\rho(3)}\|^2 + t_1t_2\|p_{\rho(2)}\|^2 + t_1\|p_{\rho(1)}\|^2$,

and the lower end points of t_1, t_2 integrals are r dependent and tend to zero as $r \to 0^+$. We shall now prove the $r \to 0^+$ limit. But first we make the following remark:

Remark 3.1: The "t_2" factorization of the quadratic form in the exponential in (3.11) is characteristic of massless Feynman amplitudes. It is in fact a special case of a more general factorization for n-point amplitudes (Lemma 2.2.20 of Ref. 12) and has implications for the structure of singularities in $\boldsymbol{\lambda}$ space. Also we have introduced imaginary parts through positive definite quadratic forms; as we shall see this enables us to prove easily the existence of the $\eta \downarrow 0$ boundary value in $\mathcal{S}'(R^{4\times3})$ for the massless amplitude.

Theorem 3.1: Let

$$\Lambda_1 = \left\{(\lambda_1,\lambda_2,\lambda_3) \mid 1 - \frac{4}{L\cdot\max_l\{\nu_l\}} < \mathrm{Re}\lambda_l < 1, 0 < \mathrm{Im}\lambda_l < a; l = 1,2,3\right\} \tag{3.13}$$

and consider

$$\boldsymbol{\lambda} \in \Lambda_1.$$

Then $\lim_{\eta\downarrow 0}\lim_{r\downarrow 0}\tilde{T}^{\nu}_{\eta,r}(\boldsymbol{\lambda})$ exists in $\mathcal{S}'(R^{4\times3})$ as a Lorentz invariant distribution. (In Sec. 3C we prove analyticity in a region $\Lambda \supset \Lambda_1$, where Λ is ν independent.)

Proof: In order to prove the theorem we study (3.8) in the sector π_ρ with the representation (3.11) for the "α integral." Noting the definition $\mu_l^\rho = \sum_{l'=1}^{l} [(\lambda_{\rho(l')} - 1)\nu_{\rho(l')}]$, the t_L integral clearly converges as $r \to 0^+$ with the regulating parameters in Λ_1. We do the t_L integral using

$$\int_0^\infty \alpha\alpha\; \alpha^{\nu-1}\exp(i\alpha\gamma) = \Gamma(\nu)\gamma^{-\nu}\, e^{(\nu\pi i)/2},$$

$$\mathrm{Im}\gamma > 0, \quad \mathrm{Re}\nu > 0$$

to get, in the $r \to 0+$ limit (which we will justify),

$$(\alpha\text{ integral})_\rho = e^{(\pi i/2)\mu_3^\rho}\,\Gamma(\mu_3^\rho + 4)$$

$$\times \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_{\rho(3)}]-1} \int_0^1 dt_1 t_1^{(\lambda_{\rho(1)}-1)\nu_{\rho(1)}+1}$$

$$\times [E(t_1t_2)]^{\mu_3^\rho+2}\cdot[P_1^\rho(\mathbf{t};\mathbf{p}) + i\eta P_2^\rho(\mathbf{t};\mathbf{p})]^{-(\mu_3^\rho+4)}. \tag{3.14}$$

First we note that $E(t_1,t_2)$ is strictly positive. The summability of the (3.14) integrand, with respect to the lower end point, for the t_1 integration follows (as for the t_3 integral) from (3.13), making use of the first inequality for $\mathrm{Re}\lambda_l$. On the other hand, the summability, with respect to lower end point of the t_2 integration, is assured by the condition $\mathrm{Re}\lambda_l < 1$ in (3.13). Hence the $r \to 0+$ limit of (3.8), i.e., of $\tilde{T}^{\nu}_{\eta,r}(\boldsymbol{\lambda})$ exists since it exists in every sector π_ρ.

We next prove the existence of the $\eta\downarrow 0$ limit of $\tilde{T}^{\nu}_{\eta,r\downarrow 0}(\boldsymbol{\lambda})$, in $\mathcal{S}'(R^{4\times3})$ using a method due to Gel'fand.[18] We smear the integrand of (3.14) [multiplied by $\delta(\sum p_i)$] with a testing function $\phi(p) \in \mathcal{S}(R^{4\times3})$ and, observing $\mathrm{Im}\mu_3 > 0$, utilize the following identity[18]:

$$\langle (P_1 + i\eta P_2)^{-\mu_3-4}, \tilde{\phi}\rangle = \{4^k[-(\mu_3+4)+1]\cdots$$

$$\times [-(\mu_3+4)+k]\cdot[-(\mu_3+4)+4]\cdots[-(\mu_3+4)$$

$$+ 4 + k - 1]\}^{-1}\cdot\langle (P_1 + i\eta P_2)^{k-\mu_3-4}, \mathcal{L}^k\tilde{\phi}\rangle, \tag{3.15}$$

where $\mathcal{L}$ is a differential operator of the form

$$\sum_{\substack{0\le\mu\le 3\\ 1\le r,s\le 2}} a_{rs}^{(\mu)} \frac{\partial}{\partial p_r^\mu}\frac{\partial}{\partial p_s^\mu},$$

$a_{rs}^{(\mu)}$ being the coefficients of the quadratic form $(P_1 + i\eta P_2)$ in p_1, p_2 only because of the application of $\delta(\sum_{i=1}^3 p_i)$.

We choose $k > 4$ and observe that $\mathrm{Re}\mu_3 < 0$. The existence of the $\eta\downarrow 0$ limit in $\mathcal{S}'(R^{4\times3})$ of (3.14), and hence of $\tilde{T}^{(\nu)}_\eta(\boldsymbol{\lambda})$, now follows since the $\eta\downarrow 0$ limit of the (smeared) t integrand of (3.14) exists and is dominated by a summable function (Lebesgue's theorem). The boundary value is clearly a Lorentz invariant distribution, since P_1 is Lorentz invariant.

Remark 3.2: We shall make one further simplification. We consider in lieu of (3.14), the expression obtained on making the replacement $P_2^\rho(\mathbf{t};\mathbf{p}) \to P_2^\rho(\mathbf{p})$, which is obtained by setting $t_j = 1,\ j = 1, 2$ in the coefficients of the quadratic form $P_2^\rho(\mathbf{t};\mathbf{p})$. Then repeating the previous chain of arguments, the boundary value $\eta \downarrow 0$ exists as a Lorentz invariant generalized function and *coincides* with the boundary value in the preceding theorem. Thus from now on we shall consider the quadratic form $[P_1(\mathbf{p},\mathbf{t}) + i\eta_3 P_2(\mathbf{p})]$ in the analytically regulated α integral (3.14): $P_2^\rho(\mathbf{p}) = \sum_{j=1} \|p_{\rho(j)}\|^2$.

(C) By virtue of Theorem 3.1 and Remark 3.2, the GFA (3.7) with $\eta, r \downarrow 0$ can be written as (in the momentum representation)

$$\tilde{T}^{\nu}_{\eta\downarrow 0}(\lambda)(\mathbf{p}) = \sum_\rho \tilde{T}^{\nu}_{\eta\downarrow 0}(\lambda)(\mathbf{p})_\rho, \qquad (3.16)$$

where

$$\tilde{T}^{\nu}_{\eta}(\lambda)(\mathbf{p})_\rho = (8\pi)^2 \delta\left(\sum_1^3 p_i\right) \exp(\tfrac{1}{2}\pi i \mu_3^\rho)$$
$$\times \prod_{l=1}^3 \left(\frac{2^{2[(\lambda_{\rho(l)}-1)\nu_{\rho(l)}]} e^{(\pi i/2)[(\lambda_{\rho(l)}-1)\nu_{\rho(l)}]}}{(4\pi^2)^{\nu_{\rho(l)}}\Gamma((1-\lambda)_{\rho(l)}\nu_{\rho(l)})}\right)$$
$$\times \Gamma(\mu_3^\rho + 4) \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_{\rho(3)}]-1}$$
$$\times A^\nu_\eta(\lambda; t_2)(\mathbf{p})_\rho \qquad (3.17)$$

and

$$A^\nu_\eta(\lambda; t_2)(\mathbf{p})_\rho = \int_0^1 dt_1\, t_1^{(\lambda_{\rho(1)}-1)\nu_{\rho(1)}+1}[E(t_1t_2)]^{\mu_3^\rho+2}$$
$$\times [P_1^\rho(\mathbf{t};\mathbf{p}) + i\eta P_2^\rho(\mathbf{p})]^{-(\mu_3^\rho+4)} \qquad (3.18)$$

with

$$\mu_3^\rho \equiv \sum_{l'=1} [(\lambda_{\rho(l')} - 1)\nu_{\rho(l')}],\quad E(t_1t_2) = 1 + t_2 + t_2t_1,$$
$$P_1^\rho(\mathbf{t},p) \equiv p_{\rho(3)}^2 + t_1t_2p_{\rho(2)}^2 + t_1p_{\rho(1)}^2,$$
$$P_2^\rho(p) \equiv \sum_{j=1}^3 \|p_{\rho(j)}\|^2,$$

with the regulating parameters in the region Λ_1 [(3.13)].

The $\eta\downarrow 0$ boundary value exists as a Lorentz invariant distribution in $\mathcal{S}'(R^{4\times 3})$.

In this subsection we will establish an analytic continuation of the GFA (3.16)–(3.18) into the ν-independent region Λ, where

$$\Lambda = \{\lambda \mid -b < \mathrm{Re}\lambda_l < 1,\quad 0 < \mathrm{Im}\lambda_l < a,\quad l = 1,2,3\},\qquad b \geq 0. \qquad (3.19)$$

Returning to (3.17) and (3.18), we observe that when

$$\mathrm{Re}\lambda_l \leq \left(1 - \frac{4}{L\cdot \max_l\{\nu_l\}}\right),$$

lower and point singularities may arise from (3.18) from the factor $t_1^{[(\lambda_\rho(1)-1)\nu_\rho(1)+1]}$; whereas the corresponding factor in (3.17) will remain integrable. The t_1 end point singularities will be reflected in complicated singularities in the regulating parameters. We first display the singularity structure.

Consider the generalized function

$$[P_1^\rho(\mathbf{p},\mathbf{t}) + i\eta P_2^\rho(p)]^{-(\mu_3^\rho+4)}, \qquad (3.20)$$

which appears in (3.18) and where P_2 is a positive definite quadratic form. Such generalized functions have been studied by Gel'fand.[18] Now we observe, following Gel'fand, that the above generalized function can be analytically continued in the coefficients of the quadratic form. In fact, *with* $\eta > 0$, it is analytic in t_1 in a complex neighborhood $K_1 \subset C'$ of the domain of integration $0 \leq t_1 \leq 1$ in (3.18). Furthermore E is analytic in t_1 and strictly positive in $0 \leq t_1 \leq 1$; it is evident that $[E(t_1t_2)]^{(\mu_3^\rho+2)}$ too is analytic in a complex neighborhood K_2 of $0 \leq t_1 \leq 1$. Let K be the common domain of extension.

Then we obtain, by a well-known procedure,[19]

$$A^\nu_\eta(\lambda; t_2)(\mathbf{p})_\rho = \frac{1}{2i\,\sin\pi[(\lambda_{\rho(1)} - 1)\nu_{\rho(1)} + 1]}$$
$$\times \int_1^{0+} dt_1 (-t_1)^{(\lambda_{\rho(1)}-1)\nu_{\rho(1)}+1}$$
$$\times [E(t_1t_2)]^{\mu_3^\rho+2}[P_1^\rho(\mathbf{t};\mathbf{p}) + i\eta P_2^\rho(\mathbf{p})]^{-(\mu_3^\rho+4)} \qquad (3.21)$$

with $(-t)^\gamma = \exp[\log|t| + i\arg(-t)]$, $|\arg(-t)| \leq \pi$. The t_1 contour begins and ends at 1, in the complex t_1 plane, encircling the origin once counterclockwise, the contour lying within K.

We remark that, from the identity of (3.21) with (3.18) in the region (3.13), the existence of the $\eta\downarrow 0$ limit in $\mathcal{S}'(R^{4\times 3})$ of (3.21) follows. [However, in (3.21), the $\eta\downarrow 0$ limit may not be interchanged with the integration.]

Formula (3.21) furnishes us with an explicit analytic continuation of (3.17) from the region Λ_1 (3.13) into the region Λ (3.19) ($\Lambda \supset \Lambda_1$) since the integral in (3.21) is analytic in λ, all the singularities being contained in the factor $\Gamma(\mu_3^\rho + 4)\cdot\{2i\,\sin\pi[(\lambda_{\rho(1)} - 1)\nu_{\rho(1)} + 1]\}^{-1}$, which is analytic in this region and the t_2 integral converges. We have introduced the representation (3.21) as it is convenient for future work.

With $\eta > 0$, (3.17) is a distribution in $\mathcal{S}'(R^{4\times 3})$ with analyticity in Λ. It now remains to prove the existence of $\eta\downarrow 0$ limit in $\mathcal{S}'(R^{4\times 3})$ of (3.17) in the region Λ. Let us return to (3.17) and (3.18) with regulating parameters in the region Λ_1. Then by repeated partial integrations[9] we get

$$A^\nu_\eta(\lambda; t_2)_\rho = \left\{\sum_{k_1=0}^{\nu_1}\left[\frac{(-1)^{k_1}(\partial/\partial t_1)^{k_1}|_{t_1=1}}{\prod_{m_1=0}^{k_1}(\mu_1 + 2 + m_1)}\right]\right.$$
$$\left. + \frac{(-1)^{\nu_1}}{\prod_{m_1=0}^{\nu_1}(\mu_1^\rho + 2 + m_1)}\int_0^1 dt_1 t_1^{\mu_1^\rho+2+\nu_1}\left(\frac{\partial}{\partial t_1}\right)^{\nu_1+1}\right\}$$
$$\times \{E(t_1t_2)^{\mu_3^\rho+2}[P_1^\rho(\mathbf{p},\mathbf{t}) + i\eta P_2^\rho(\mathbf{p})]^{-(\mu_3^\rho+4)}\} \qquad (3.22)$$

(with $\mu_l^\rho = \sum_1^l[(\lambda_{\rho(l')} - 1)\nu_{\rho(l')}]$), which may be expanded cut as a sum of terms. This formula with all end point singularities factored out also provides us with an explicit analytic continuation of (3.17) to the region Λ. From the theorem on analytic continuations, it follows that, in the region Λ, (3.21) and (3.22) [and hence the corresponding representations for (3.17)] are *identical*. Smear $A^\nu_\eta(\lambda; t_2)$ multiplied by $\delta(\sum p)$ with a test function $\tilde{\phi}(p) \in \mathcal{S}(R^{4\times 3})$ and apply again Gel'fand's method [to the representation (3.22)] as in Sec. 3B. Note that the application of various derivatives in (3.22) will lower the exponent of the quadratic form [last factor in (3.22)] by a finite (ν_1 dependent) amount. On making use again of the identity (3.15), we can raise the power up so that the real part of the exponent is *positive*. Then

$\langle (P_1 + i\eta P_2)^{-(\mu_3^\rho+4)}, \tilde{\phi}\rangle$ is *continuous* in η. The existence of the $\eta \downarrow 0$ limit in $\mathcal{S}'(R^{4\times 3})$ of $\delta(\sum p) A_\eta^{\boldsymbol{\nu}}(\lambda; t_2)$ (in Λ), and hence of (3.17) now follows by dominated convergence. (Note that the boundary value is a Lorentz invariant distribution since P_1 is Lorentz invariant.)

We have now obtained the definition of our regulated massless Feynman amplitude (3.17) [the regulated form of the formal product (3.3)] as a distribution with regulating parameters in the region of interest (3.19). We state this result as a theorem.

Theorem 3.2: The generalized (massless) Feynman amplitude [corresponding to the formal product (3.3)] is given by the Fourier transform of

$$\tilde{\mathcal{T}}^{\boldsymbol{\nu}}(\lambda)(\mathbf{p}) = \sum_\rho \tilde{\mathcal{T}}^{\boldsymbol{\nu}}_{\eta\downarrow 0}(\lambda)(\mathbf{p})_\rho \tag{3.23}$$

with

$$\tilde{\mathcal{T}}^{\boldsymbol{\nu}}_\eta(\boldsymbol{\lambda})(\mathbf{p})_\rho = (8\pi)^2 \delta\left(\sum_1^3 p_i\right) \exp(\tfrac{1}{2}\pi i\, \mu_3^\rho) \prod_{l=1}^3 \times \left(\frac{2^{2(\lambda_{\rho(l)}-1)\nu_{\rho(l)}}}{\Gamma((1-\lambda_{\rho(l)})\nu_{\rho(l)})} \frac{e^{(i\pi/2)(\lambda_{\rho(l)}-1)\nu_{\rho(l)}}}{(4\pi^2)^{\nu_{\rho(l)}}}\right)\Gamma(\mu_3^\rho + 4) \times \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_{\rho(3)}]-1} A_\eta^{\boldsymbol{\nu}}(\lambda; t_2)(\mathbf{p})_\rho \tag{3.24}$$

and

$$A_\eta^{\boldsymbol{\nu}}(\boldsymbol{\lambda}; t_2) = \{2i \sin\pi[(\lambda_{\rho(1)} - 1)\nu_{\rho(1)} + 1]\}^{-1} \times \int_1^{0+} dt_1 (-t_1)^{(\lambda_{\rho(1)}-1)\nu_{\rho(1)}+1} \times [E(t_1 t_2)]^{\mu_3^\rho+2} [P_1^\rho(\mathbf{t};\mathbf{p}) + i\eta P_2^\rho(\mathbf{p})]^{-(\mu_3^\rho+4)}, \tag{3.25}$$

where $\mu_3^\rho = \sum_1^3 (\lambda_{\rho(l)} - 1)\nu_{\rho(l)}$ and $\lambda_1 \cdots \lambda_3$ are restricted to the region

$$\Lambda = \{\boldsymbol{\lambda} \mid -b < \mathrm{Re}\lambda_l < 1, 0 < \mathrm{Im}\lambda_l < a, l = 1,2,3\}, \quad b \geq 0. \tag{3.26}$$

Then (3.23) defines a Lorentz invariant distribution in $\mathcal{S}'(R^{4\times 3})$ with analyticity in Λ and meromorphic elsewhere. For fixed $\nu_1 \cdots \nu_3$ (taken arbitrarily) the singularity of $\mathcal{T}^{\boldsymbol{\nu}}_{\eta\downarrow 0}(\lambda)$ at $\boldsymbol{\lambda} = \mathbf{0}$ is in the form of poles on hyperplanes in $\mathbb{C}^3$:

$$(\lambda_{\rho(1)} - 1)\nu_{\rho(1)} = -k_1 - 2, \quad \sum_1^3 (\lambda_{\rho(l)} - 1)\nu_{\rho(l)} = -k_2 - 4$$

with $K_j = 0, 1, 2, \cdots$ $(j = 1, 2)$. [This is evident from the factors $\Gamma(\mu_3^\rho + 4)$ in (3.24) and $\{2i \sin\pi[(\lambda_{\rho(1)} - 1) \times \nu_{\rho(1)} + 1]\}^{-1}$ in (3.25). The other poles from this factor in (3.25) are spurious since their residues vanish.]

Remark 3.3: We have given the construction of the analytically regulated Feynman amplitude with massless multiplets for the three-point function, since it is with this case that we shall be concerned in this paper. Our procedure readily generalizes[20] for the n-point function on making appropriate use of the general $\mathbf{t}$ factorization mentioned in Remark 3.1 and will be given in a different context.

(D) We are now in a position to turn to our basic object of concern: the analytically regulated version of $\mathcal{T}(\mathbf{x})$ given in (3.2). Let us define

$$\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})(\mathbf{p}) = \sum_{\nu_1,\nu_2,\nu_3=1}^\infty \frac{(g^2)^{\sum_1^3(1-\lambda_l)\nu_l}}{\prod_{l=1}^3[\nu_l!]} \tilde{\mathcal{T}}^{\boldsymbol{\nu}}_\eta(\boldsymbol{\lambda})(\mathbf{p}), \tag{3.27}$$

where the GFA $\tilde{\mathcal{T}}^{\boldsymbol{\nu}}_\eta$ is given through (3.23)–(3.25) and the $\boldsymbol{\lambda}$ regulators are restricted to the region Λ given in (3.26). By Theorem 3.2, each term in (3.27) (together with the $\eta\downarrow 0$ limit) is a distribution in $\mathcal{S}'(R^{4\times 3})$. Let us smear each term in (3.27) with a test function $\tilde{\phi}(\mathbf{p}) \in \mathfrak{M}_g(R^{4\times 3})$. Returning to (3.24), it is permissible to smear $A_\eta^{\boldsymbol{\nu}}(\lambda), t_2)$ times $\delta(\sum p)$. In Appendix A we show that, for large $\|\nu\|_3 = \sum_1^3 \nu_l$,

$$|\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| \leq L_\eta \tilde{\psi}_\eta^{(\nu)} \|\tilde{\phi}\|_g \|\nu\|_3^{[(1/\rho)\|\nu\|_3)]} \tag{3.28}$$

and that a similar bound holds in the limit $\eta\downarrow 0$. The $\delta(\sum p)$ restriction is understood. Here $\tilde{\psi}^{(\nu)}$ has growth (const$^{\|\nu\|_3}$ $\|\tilde{\phi}\|_g$ is a norm in $\mathfrak{M}_g(R^{4\times 2})$ and ρ is the order of growth of the indicator function $g(t)$ (Ref. 1). Choosing $\rho > \frac{1}{3}$, as in Sec. 2, and using Stirling's formula, we verify that (3.27) converges in the topology of $\mathfrak{M}'_g(R^{4\times 3})$ and uniformly in η. Hence $\tilde{\mathcal{T}}_{\eta\downarrow 0}(\lambda)$ exists in $\mathfrak{M}'_g(R^{4\times 3})$ and is analytic in Λ, because the convergence is uniform in compact subsets of Λ. We thus have the following:

Theorem 3.3:

$$\mathcal{T}_{\eta\downarrow 0}(\boldsymbol{\lambda})(\mathbf{x}) = \sum_{\nu_1\cdots\nu_3=1}^\infty \frac{(g^2)^{\sum_1^3(1-\lambda_l)\nu_l}}{\prod_1^3[\nu_l!]} \mathcal{T}_{\eta\downarrow 0}(\boldsymbol{\lambda})(\mathbf{x}) \tag{3.29}$$

exists as a Lorentz invariant (strictly localizable) generalized function in a space $\mathfrak{M}'_g(R^{4\times 3})$, with indicator function of order of growth ρ, $\frac{1}{3} < \rho < \frac{1}{2}$, analytic in the region Λ [(3.26)].

(3.29) constitutes the analytically regulated version of the formal series (3.2) of divergent Feynman graphs. For later purposes it is convenient to express (3.27), via (3.23)–(3.25), as

$$\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})(\mathbf{p}) = (8\pi)^2 \delta\left(\sum_1^3 p_i\right) \sum_{\nu_1\cdots\nu_3=1}^\infty \prod_1^3 \left(\frac{1}{[(4\pi)^2]^{\nu_l}\Gamma(\nu_l+1)}\right) \times \sum_\rho \prod_1^3 \left(\frac{2^{2\lambda_{\rho(l)}\nu_l} e^{(\lambda_{\rho(l)}-1)\nu_l \pi i}}{\Gamma((1-\lambda)_{\rho(l)}\nu_l)} (g^2)^{(1-\lambda_{\rho(l)})\nu_l}\right) \times \Gamma(\mu_3^\rho + 4) \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_3]-1} A_\eta^{\boldsymbol{\nu}}(\boldsymbol{\lambda}; t_2)(\mathbf{p})_\rho \tag{3.30}$$

and

$$A_\eta^{\boldsymbol{\nu}}(\boldsymbol{\lambda}; t_2) = \{2i \sin\pi[(\lambda_{\rho(1)} - 1)\nu_1 + 1]\}^{-1} \times \int_1^{0+} dt_1 (-t_1)^{(\lambda_{\rho(1)}-1)\nu_1+1} \times [E(t_1 t_2)]^{\mu_3^\rho+2} [P_1^\rho(\mathbf{p}, t) + i\eta P_2^\rho(\mathbf{p})]^{-(\mu_3^\rho+4)} \tag{3.31}$$

with

$$\mu_3^\rho = \sum_1^3 (\lambda_{\rho(l)} - 1)\nu_l.$$

In order to derive the above expressions we have exploited the fact that the series coefficient satisfy

$$\frac{(g^2)^{\sum_1^3(1-\lambda_l)\nu_l}}{\prod_1^3\{[(4\pi)^2]^{\nu_l}\Gamma(\nu_l+1)\}} = \frac{(g^2)^{\sum_1^3(1-\lambda_{\rho(l)})\nu_{\rho(l)}}}{\prod_1^3\{[(4\pi)^2]^{\nu_{\rho(l)}}\Gamma(\nu_{\rho(l)}+1)\}}$$

for any permutation ρ of $(1,2,3)$. We have then interchanged $\sum_\rho$ with $\sum_{\{\nu_i\}}$ which is then the same as $\sum_{\{\nu_{\rho(i)}\}}$. Since the $\nu_{\rho(i)}$ are summed over, we can replace $\nu_{\rho(i)} \to \nu_i$ to get the above expression.

4. RENORMALIZATION OF THE THREE POINT FUNCTION

(A) In this section we will show how the series of regulated Feynman graphs (3.29) may be defined at

$\lambda = 0$ through a renormalization. We already remarked (Theorem 3.2) on the singularity structure of the individual GFA's $\tilde{\mathcal{T}}^{\nu}_{\eta}(\lambda)$ for arbitrary (but fixed) ν. For an individual GFA, one could renormalize, by procedures suitable for meromorphic functions either through a generalized evaluation[9] or through an analytic evaluation[10] which is equivalent to an analytic decomposition into regular and singular parts and analytic continuation of the regular part to $\lambda = 0$. However, $\tilde{\mathcal{T}}_{\eta}(\lambda)$ being the sum of an infinite series of GFA's is not meromorphic, by the reasoning of Sec. 2. We shall therefore generalize the extension procedure of Sec. 2 in order to give an analytic decomposition of $\tilde{\mathcal{T}}_{\eta}(\lambda)$ into a sum of regular and singular parts.

Returning to (3.30) and (3.31), it is convenient to write the ν_1 sum as a Z contour integral representation (as in Sec. 2), since (as will be clear) the λ dependent poles on hyperplanes in $\mathbb{C}^3$ can be analysed as moving Z plane poles in the spirit of Sec. 2. We get

$$\tilde{\mathcal{T}}_{\eta}(\lambda)(\mathbf{p}) = (8\pi)^2\, \delta\left(\sum_{i=1}^{3} p_i\right) \sum_{\nu_2,\nu_3=1}^{\infty} \prod_{l=2}^{3} \left(\frac{1}{\Gamma(\nu_l + 1)}\right) \times \sum_{\rho} \tilde{F}^{\nu}_{\eta}(\lambda)(\mathbf{p})_{\rho}, \qquad (4.1)$$

where

$$\tilde{F}^{\nu}_{\eta}(\lambda)(\mathbf{p})_{\rho} = \prod_{l=2}^{3} \left(2^{2\lambda_{\rho(l)}\nu_l} \frac{e^{(\lambda_{\rho(l)}-1)\nu_l \pi i}}{\Gamma((1-\lambda_{\rho(l)})\nu_l)} (g^2)^{(1-\lambda_{\rho(l)})\nu_l}\right) \times \frac{1}{2i} \int_{\Gamma_1} dz\, \frac{2^{2\lambda_{\rho(1)}z} e^{(\lambda_{\rho(1)}-1)z\pi} (g^2)^{(1-\lambda_{\rho(1)})z}}{[4\pi)^2]^z\, \Gamma((1-\lambda_{\rho(1)})z)\Gamma(z+1)}$$
$$\times \cot\pi z\, \Gamma(\mu_3^{\rho} + 4) \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_3]-1} A^{\nu}_{\eta}(z, \lambda; t_2)(\mathbf{p})_{\rho} \qquad (4.2)$$

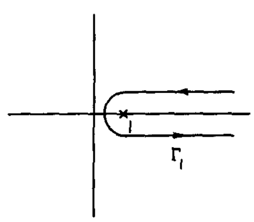

FIG. 4. The contour Γ_1.

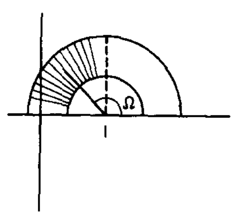

FIG. 5. The shaded region Λ^{l}.

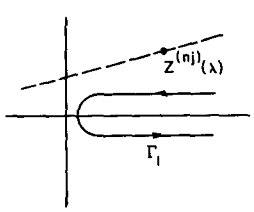

FIG. 6. Location of moving poles.

with A^{ν}_{η} given as in (3.31) with the replacement $\nu_1 \to z$ and

$$\mu_3^{\rho} \to \mu_3^{\rho} = (\lambda_{\rho(1)} - 1)z + \sum_{l=2}^{3} (\lambda_{\rho(l)} - 1)\nu_l .$$

The contour Γ_1 is shown in Fig. 4. From the following analysis it will follow that Γ_1 can be chosen so as to enclose only the fixed poles of $\cot\pi z$ at the positive integer points: (4.1) when smeared with

Now $\tilde{F}^{\nu}_{\eta}(\lambda)$ has *possible* singularities at the nonanalytic points of the integrand of (4.2). Aside from the fixed poles of $\cot\pi z$, these are moving (λ-dependent) singularities in the z plane; they lead to singularities of $\tilde{F}$ (as $\lambda \to 0$) only if they cannot be avoided by Γ_1 contour distortions.

The Γ function in (4.2) and the over-all factor (in braces) in the definition (3.31) of A^{ν}_{η} (in 4.2) lead to the following set of moving z-plane poles:

$\tilde{\phi} \in \mathfrak{M}_g(R^{4\times 3})$ converges and defines a generalized function in $\mathfrak{M}'_g(R^{4\times 3})$. Let us now see how the $\boldsymbol{\lambda}$ singularities arise.

First, consider a subregion from (3.26) in $\mathbb{C}^3$:

$$\Lambda_s = \Lambda^{(1)} \times \Lambda^{(2)} \times \Lambda^{(3)},$$
$$\Lambda^{(l)} = \{\lambda_l \mid \lambda_l - 1 = r_l e^{i\Omega_l}, \tfrac{1}{2}\pi < \Omega_l < \pi, |r_l - 1| < \epsilon_l\} \qquad (4.3)$$

is the shaded region in Fig. 5.

$$z^{(n_1)}(\lambda) = (n_1 + 2)/(1 - \lambda_{\rho(1)}),$$
$$z^{(n_3)}(\lambda) = \left(n_3 + 4 - \sum_{2}^{3} \frac{(1-\lambda_{\rho(l)})\nu_l}{1 - \lambda_{\rho(l)}}\right), \qquad (4.4)$$

with $n_j = 0, 1, 2, \cdots$ ($j = 1, 3$) and all permutations ρ. From (4.3) and (4.4) we get (with the convention $\sum = 0$ for $j = 1$, $n_1^0 = 2$, $n_3^0 = 4$)

$$\mathrm{Re}\, z^{(n_j)}(\lambda) = \frac{1}{r_{\rho(1)}} [(n_j + n_j^0) \cos\Omega_{\rho(1)} + \sum_{l=z}^{j} r_{\rho(l)} \nu_l \cos(\Omega_{\rho(l)} - \Omega_{\rho(1)})], \qquad (4.5)$$

$$\mathrm{Im}\, z^{(n_j)}(\lambda) = \frac{1}{r_{\rho(1)}} \left((n_j + n_j^0) \sin\Omega_{\rho(1)} - \sum_{2}^{j} r_{\rho(l)} \nu_l \sin(\Omega_{\rho(l)} - \Omega_{\rho(1)})\right),$$

and hence

$$\mathrm{Im}\, z^{(n_j)} = -\tan\Omega_{\rho(1)}\, \mathrm{Re}\, z^{(n_j)} - \frac{1}{r_{\rho(1)} \cos\Omega_{\rho(1)}} \sum_{l=2}^{j} r_{\rho(l)} \nu_l \sin\Omega_{\rho(l)}. \qquad (4.6)$$

From (4.5) and (4.6) we see that the $z^{n_j}(\lambda)$ poles with $\mathrm{Re}\, z^{n_j} \geq 0$ lie on a straight line (in the z plane) with a positive intercept and subtend an angle $\Omega'_{\rho(1)}$ $(= \pi - \Omega_{\rho(1)})$, $0 < \Omega'_{\rho(1)} < \pi/2$ on the real axis. Clearly, for $\mathrm{Re}\, z^{n_j} > 0$, we have $\mathrm{Im}\, z^{n_j} > 0$. Thus, as shown in Fig. 6, the Γ_1 contour can be chosen so that all $z^{n_j}(\lambda)$ poles lie exterior to it.

In the region Λ_s (4.3), (4.1), and (4.2) are analytic in λ. However, if we were to continue to $\lambda \to 0$, the Γ_1 contour would get pinched between the fixed poles at $z = k$ ($k = 1, 2, 3, \cdots$) and, from (4.4), those moving poles $z^{n_j}(\boldsymbol{\lambda})$ which at $\boldsymbol{\lambda} = 0$ satisfy

$$\begin{aligned} z^{n_1} &= n_1 + 2 = k_1, \\ z^{n_3} &= n_3 + 4 - \sum_2^3 \nu_l = k_3, \quad k_j = 1, 2, 3, \cdots, \end{aligned} \tag{4.7}$$

leading to nonanalyticity of (4.1) and (4.2) at $\boldsymbol{\lambda} = 0$.

The first set of poles in (4.4) is identical to the set in Sec. 2; following the procedure of that section, we distort the contour $\Gamma_1 \to \overline{\Gamma}_1$ so as to enclose the $z^{(n_1)}(\lambda)$ poles and compute the discontinuities. This leads to the decomposition

$$\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda}) = \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})_{R_1} + \Delta_1 \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda}) \tag{4.8}$$

of (4.1) in Λ_δ. We remark that holding any two λ_j, λ_k fixed in $\Lambda^{(j)}, \Lambda^{(k)}$, we have that (i) $\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})_{R_1}$ is analytic in $\lambda_l \in \Lambda^{(l)}$ and (ii) $\lim_{\lambda_l \to 0} \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})_{R_1}$ exists uniquely for any sequence $\{\lambda_l^i \mid \lambda_l^i \in \Lambda^{(l)}\}$ since it may be seen that the $z^{n_j}(\boldsymbol{\lambda})$ poles do not lead to pinching of $\overline{\Gamma}_1$. Next we consider the second set of poles in (4.4) which will cause a singularity of $\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})_{R_1}$ when all $\lambda_l = 0$ $(l = 1, 2, 3)$. We set $\Omega_l = \Omega$ and deduce from (4.5) and (4.6), for $\operatorname{Re} z^{n_3} > 0$:

$$0 < \operatorname{Im} z^{n_3}(\lambda) = \tan\Omega' \operatorname{Re} z^{(n_3)}(\lambda) + \tan\Omega'(\nu_2 + \nu_3), \tag{4.9}$$

where $\Omega' = \pi - \Omega$.

Returning to (4.8), we repeat the contour distortion procedure $\Gamma_1 \to \overline{\Gamma}$ so as to enclose the $z^{(n_3)}(\lambda)$ poles with $(\operatorname{Re} z^{n_3} \geq 1)$ and compute the discontinuity (see Figs. 7 and 8). Because of (4.9), in order to enclose $z^{n_3}(\lambda)$ poles, it is sufficient at this stage to have the upper edge of the contour $\overline{\Gamma}$ to be a piece of the straight line subtending an angle Ω' on the real axis and an intercept $(\nu_2 + \nu_3 + \epsilon)\tan\Omega'$, $\epsilon > 0$, on the imaginary axis (it may be checked that such a distortion does not lead to any divergence).

We then get the analytic decomposition in Λ:

$$\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda}) = \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})_R + \Delta_1 \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda}) + \Delta_3 \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda}). \tag{4.10}$$

Now $\Delta_{1,3}\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})$ (which are the residues of the moving poles) have a complicated singularity at $\boldsymbol{\lambda} = 0$. However, as shown in Appendix B, $\mathcal{F}^{-1}[\Delta_3 \tilde{\mathcal{T}}_{\eta\downarrow 0}(\boldsymbol{\lambda})](\mathbf{x})$ is a distribution in $\mathcal{C}_g'(R^{4\times 3})$ with support concentrated on $x_1 = x_2 = x_3$. Furthermore, $\mathcal{F}^{-1}[\Delta_1 \tilde{\mathcal{T}}_{\eta\downarrow 0}(\boldsymbol{\lambda})](\mathbf{x})$ is a sum of distributions with supports concentrated on surfaces $x_i = x_j$. On the other hand, $\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})_R$ can be analytically continued to $\boldsymbol{\lambda} = 0$. This continuation may be done by setting all the λ's equal to some λ and continuing in the single variable to zero. (The analyticity follows from the fact that in the continuation, no further $\overline{\Gamma}$ contour distortion is necessary; the moving $z^{n_j}(\boldsymbol{\lambda})$ poles fusing with the fixed poles at the positive integers.) It now follows that $\mathcal{F}^{-1}[\tilde{\mathcal{T}}_{\eta\downarrow 0}(\boldsymbol{\lambda})](x)$ is unambiguously defined as a continuous linear functional only on the subspace $\mathcal{C}_0(R^{4\times 3})$ of test functions from $\mathcal{C}_g(R^{4\times 3})$ which vanish, together with all derivatives, when any two space-time points coincide. This was to be expected.

We now obtain a definition of $\mathcal{T}(\boldsymbol{\lambda})(\mathbf{x})$ at $\boldsymbol{\lambda} = 0$ through the following extension to $\mathcal{C}_g'(R^{4\times 3})$:

$$\mathcal{T}_3(\mathbf{x}) = \mathcal{F}^{-1}(\tilde{\mathcal{T}}_{\eta\downarrow 0})(\mathbf{x}), \tag{4.11}$$

where

$$\begin{aligned} \tilde{\mathcal{T}}_\eta(\mathbf{p}) &= \lim_{\boldsymbol{\lambda}\to 0} \tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})(\mathbf{p})_R \\ &= \delta\Big(\sum_{i=1}^3 p_i\Big) \sum_{\nu_2,\nu_3=1}^{\infty} \prod_{l=2}^{3} \left(\frac{(-1)^{\nu_l}(g^2)^{\nu_l}}{[(4\pi)^2]^{\nu_l}\Gamma(\nu_l+1)}\right) \frac{1}{2i} \\ &\quad \times \int_{\overline{\Gamma}} dz \, \frac{e^{-\pi i z}\cot\pi z\,(g^2)^z}{[(4\pi)^2]^z\Gamma(z)\Gamma(z+1)} \, \frac{\Gamma(4 - z - \Sigma_2^3 \nu_l)}{2i\sin\pi(-z+1)} \\ &\quad \times \int_0^1 dt_2\, t_2^{\nu_3 - 1} \int_1^{0+} dt_1 (-t_1)^{-z+1} [E(t_1 t_2)]^{2 - z - \Sigma_2^3 \nu_l} \\ &\quad \times \sum_\rho [P_1^\rho(\mathbf{t}, \mathbf{p}) + i\eta P_2^\rho(\mathbf{p})]^{z + \Sigma_2^3 \nu_l - 4}. \end{aligned} \tag{4.12}$$

From the considerations of Sec. 3D, it follows that $\tilde{\mathcal{T}}_{\eta\downarrow 0}$ is a Lorentz invariant distribution in $\mathfrak{M}'(R^{4\times 3})$. The extension (4.11) and (4.12) is certainly not unique. We can clearly add to the rhs of (4.11) any translation invariant distribution in $\mathcal{C}_g'(R^{4\times 3})$, $\mathfrak{X}(x_1, x_2, x_3)$ with support concentrated on $x_1 = x_2 = x_3$. In fact we shall take advantage of this later to obtain a minimal and unique extension. But first we prove that the extension (4.11) and (4.12) constitutes a renormalization.

(B) We can characterize a renormalization through a certain number of properties.[21] Basic to it is the property of being a Lorentz invariant extension to a distribution (in $\mathcal{C}'$ for us); this we have already shown. Furthermore, the extension of $\overline{\mathcal{T}}(\mathbf{x})$ is the complex conjugate of the extension $\mathcal{T}(\mathbf{x})$. [This follows easily on utilizing the fact that in Sec. 3 the $\overline{\mathcal{T}}^\nu(\boldsymbol{\lambda})$ can be obtained from $\mathcal{T}^\nu(\boldsymbol{\lambda})$ by complex conjugation for real values of the regulating parameters, and then by analytic continuation in general.] It remains to prove the basic unitarity—causality relations for the functions.

Starting from the unitarity—causality relations for T products,[2,3] and the relation (3.1), a straightforward application of Wick's theorem leads to the following sufficiency conditions for the respective properties of the T products:

Unitarity:

$$0 = \sum_{U \subset G} (-1)^{|U|} \, \overline{\mathcal{T}}_{G(U)}$$

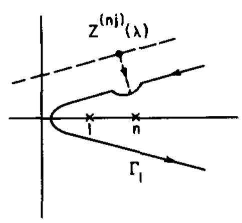

FIG. 7. Pinching of Γ_1.

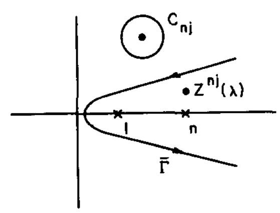

FIG. 8. Contour splitting.

$$\times \sum_{\substack{\nu_l=1\\ l\in M}}^{\infty} \prod_{l\in M(U,U')} \left(\frac{(g^2)^{\nu_l}\Delta_+^{\nu_l}(v_{il},v_{fl})}{\nu_l!}\right) \mathcal{T}_{G(U')}. \quad (4.13)$$

Here $U = \{V_{j1}\ldots, V_{jk}\}$ is a generalized vertex of $G(V_1V_2V_3,), U' = C(U), \mathcal{T}_{G(U)} = \mathcal{T}(x_{jl},\cdots,x_{jk})$ and $M(U,U')$ is the set of lines connecting U and U'. For notational simplicity we will suppress the space–time dependence.

Causality:

$$\operatorname{supp} R_3(x_1x_2x_3) \subset \{x_1 - x_j \in \overline{V}_+, j = 2,3\}, \quad (4.14)$$

where

$$R_3 \equiv \sum_{U_1\subset G} (-1)^{|U_1|-1}\, \overline{\mathcal{T}}_{G(U_1)} \sum_{\substack{\nu_l=1\\ l\in M}}^{\infty} \left(\prod_{l\in M(U_1,U_1')} \frac{(g^2)^{\nu_l}\Delta_+^{\nu_l}(v_{il},v_{fl})}{\nu_l!}\right) \mathcal{T}_{G(U_1')} \quad (4.15)$$

where $U_1 = U$ such that $V_1 \in U$. In order to show that the $\mathcal{T}_3$ functions of (4.11) and (4.12) ($\mathcal{T}_2$ was obtained in Sec. 2) satisfy the above conditions, we first obtain an analytically regulated version[22] of these relations.

First we note that the work of Section 3B could have been done using Pauli–Villars (p.v.) regulators $\{M_j\}$ [22] instead of the regulator "r". Then the analytically and p.v. regulated propagator replacing (3.7) would be proportional to

$$\lim_{\eta\downarrow 0} \int_0^{\infty} d\alpha_l\, \alpha_l^{[(\lambda_l-1)\nu_l]+1} I(\alpha_l,\mathbf{M}) \exp i\alpha_l[p^2 + i\eta \,\|p\|^2], \quad (4.16)$$

where, for sufficiently many regulator masses, $I(\alpha,\mathbf{M})$ can have a zero of any desirable order at $\alpha_l = 0$. Then the identities

$$\Delta^{(\nu),M}(\lambda;x) = \theta(x^0)\Delta_+^{(\nu),M}(\lambda;x) + \theta(-x^0)\Delta_+^{(\nu),M}(\lambda;-x),$$
$$\overline{\Delta}^{(\nu),M}(\lambda;x) = \theta(-x^0)\Delta_+^{(\nu),M}(\lambda;x) + \theta(x^0)\Delta_+^{(\nu),M}(\lambda;-x) \quad (4.17)$$

hold in the sense of continuous functions, and the graphwise unitarity relation

$$0 = \sum_{U\subset G} (-1)^{|U|} \Big[\prod_{l\in\mathcal{L}[G(U)]} \overline{\Delta}^{(\nu_l),M}(\lambda_l)\Big] \times \Big[\prod_{l\in M(U,U')} \Delta_+^{(\nu_l),M}(\lambda_l)\Big] \Big[\prod_{l\in\mathcal{L}[G(U')]} \Delta^{(\nu_l),M}(\lambda_l)\Big] \quad (4.18)$$

holds as an algebraic identity between continuous functions.[22] Here $\mathcal{L}[G(U)]$ is the set of lines connecting vertices of U. Recall from Sec. 3 that each "line" is a multiplet of ν_l lines of the Feynman graph. The first and third factors can be evaluated in momentum space, as in 3B, and from Theorem 3.2 it follows that the limit $M_j \to \infty$ exists if we restrict $\boldsymbol{\lambda}$ to the region Λ_1. We then get (with the p.v. regulators removed) the generalized unitarity relation for the massless GFA $\mathcal{T}_3^{\boldsymbol{\nu}}(\boldsymbol{\lambda})(\mathbf{x})$ in the region Λ_1. Now we can carry through the analytic continuation procedure of Sec. 3C leading to Theorem 3.2 [note that $\Delta_+^{(\nu_l)}(\lambda_l)$ is entire analytic] and obtain the unitarity relation for the massless GFA's:

$$0 = \sum_{U\subset G} (-1)^{|U|}\, \overline{\mathcal{T}}^{\boldsymbol{\nu}}_{G(U)}(\boldsymbol{\lambda})\Big\{\prod_{l\in M(U,U')} \Delta_+^{\nu_l}(\lambda_l)\,[\mathcal{T}^{\boldsymbol{\nu}}_{G(U')}(\boldsymbol{\lambda})]\Big\} \quad (4.19)$$

with the λ's (which are distinct for each factor) restricted to the region Λ [(3.26)]. We multiply (4.19) by

$$(g^2)^{\sum_1^3(1-\lambda_l)\nu_l} \prod_{l\in\mathcal{L}(G)} \left(\frac{1}{\Gamma(\nu_l+1)}\right) \quad (4.20)$$

and redistribute it among the various factors in (4.19). We then get

$$0 = \sum_{U\subset G} (-1)^{|U|}\left[\left(\prod_{l\in\mathcal{L}[G(U)]} \frac{(g^2)^{(1-\lambda_l)\nu_l}}{\Gamma(\nu_l+1)}\right) \overline{\mathcal{T}}^{\boldsymbol{\nu}}_{G(U)}(\boldsymbol{\lambda})\right] \times \prod_{l\in M(U,U')}\left(\frac{(g^2)^{(1-\lambda_l)\nu_l}\Delta_+^{(\nu_l)}(\lambda_l)}{\Gamma(\nu_l+1)}\right) \times \left[\left(\prod_{l\in\mathcal{L}[G(U')]} \frac{(g^2)^{(1-\lambda_l)\nu_l}}{\Gamma(\nu_l+1)}\right)\mathcal{T}^{\boldsymbol{\nu}}_{G(U')}(\boldsymbol{\lambda})\right]. \quad (4.21)$$

Finally, summing over all the ν we get, noting (3.29) the generalized unitarity relation,

$$0 = \mathcal{T}_3(\boldsymbol{\lambda}) - \overline{\mathcal{T}}_3(\boldsymbol{\lambda}) + \sum_{\substack{U\subset G\\ U\neq G,\neq\phi}} (-1)^{|U|}\overline{\mathcal{T}}_{G(U)}(\boldsymbol{\lambda}) \times \sum_{\substack{\nu_l=1\\ l\in M}} \left(\prod_{l\in M(U,U)} \frac{(g^2)^{(1-\lambda_l)\nu_l}\Delta_+^{\nu_l}(\lambda_l)}{\Gamma(\nu_l+1)}\right)\mathcal{T}_{G(U')}(\boldsymbol{\lambda}) \quad (4.22)$$

with the λ's in the region Λ. A generalized causality relation for the analytically regulated $\mathcal{T}$ functions can be similarly derived.

We can apply to (4.22) the analytic decomposition method of Sec. 2 and Sec. 4A in the subregion Λ_s of (4.3). Considering the $Z^{n_1}(\boldsymbol{\lambda})$ poles of (4.4) and Sec. 2, we get the decomposition (2.7) and (4.8). Hence, (4.22) can be written as

$$0 = \Bigg[\mathcal{T}_3(\boldsymbol{\lambda})_{R_1} - \overline{\mathcal{T}}_3(\boldsymbol{\lambda})_{R_1} + \sum_{\substack{U\subset G\\ U\neq G,\neq\phi}} (-1)^{|U|}\,\overline{\mathcal{T}}_{G(U)}(\boldsymbol{\lambda})_R \times \sum_{\substack{\nu_l=1\\ l\in M}}^{\infty}\left(\prod_{l\in M(U,U')} \frac{(g^2)^{(1-\lambda_l)\nu_l}\Delta_+^{\nu_l}(\lambda_l)}{\Gamma(\nu_l+1)}\right)\cdot \mathcal{T}_{G(U')}(\boldsymbol{\lambda})_R\Bigg] + \Delta_1 U(\boldsymbol{\lambda}), \quad (4.23)$$

where

$$\Delta_1 U(\boldsymbol{\lambda})(\mathbf{x}) = \sum_{\rho}\sum_{n=2}^{\infty} a_n^{\rho}(\lambda_{\rho(1)}) \times \sum_{r=0}^{n-2} f_{r,n}^{\rho}(\lambda_{\rho(2)},\lambda_{\rho(3)})\,\Box^{r}\delta(x_{\rho(1)_i} - x_{\rho(1)_f}) \quad (4.24)$$

as follows from (2.8) and Appendix B (B4). The identifications can be easily made. It suffices to note that $a_n^{\rho}(\lambda_{\rho(1)})$ is proportional to $\cot[\pi n/(1-\lambda_{\rho(1)})]$ times analytic factors with no zeroes for $\lambda_{\rho(1)} = 0$. On the other hand, $\lim f_{r,n}^{\rho}(\lambda_{\rho(2)},\lambda_{\rho(3)})$ as $\lambda_{\rho(2)} \to 0$ exists for fixed $\lambda_{\rho(3)}$ and similarly for the other variable. Now we show $\Delta_1 U(\boldsymbol{\lambda})(\mathbf{x}) = 0$. Fixing our attention on $\lambda_{\rho(1)}$, for some ρ, we recall [from Sec. 2 and the remark after (4.8)] that the limit $\lambda_{\rho(1)} \to 0$ (for fixed $\lambda_{\rho(2)},\lambda_{\rho(3)}$) exists for the square bracket in (4.23). Hence, under the same conditions, $\lim \Delta_1 U(\boldsymbol{\lambda})(\mathbf{x})$ as $\lambda_{\rho(1)} \to 0$ exists. On the other hand, (4.24) can be written as

$$\Delta_1 U(\boldsymbol{\lambda})(\mathbf{x}) = \sum_{\rho}\sum_{r=0}^{\infty} b_r^{\rho}(\boldsymbol{\lambda})\,\Box^{r}\delta(x_{\rho(1)_i} - x_{\rho(1)_f}), \quad (4.25)$$

where

$$b_r^\rho(\boldsymbol{\lambda}) = \sum_{n=r}^{\infty} a_{n+2}^\rho(\lambda_{\rho(1)}) f_{r,n+2}^\rho(\lambda_{\rho(2)}, \lambda_{\rho(3)})$$

and we must have $\lim b_r^\rho(\boldsymbol{\lambda})$ as $\boldsymbol{\lambda}_{\rho(1)} \to 0$ exists for each r. But

$$b_r^\rho(\boldsymbol{\lambda}) - b_{r+1}^\rho(\boldsymbol{\lambda}) = a_{r+2}^\rho(\lambda_{\rho(1)}) f_{r,r+2}^\rho,$$

where $a_{r+2}^\rho(\lambda_{\rho(1)})$ has a pole at $\lambda_{\rho(1)} = 0$. Hence $f_{r,r+2} = 0$. By a trivial induction we get for each r, $f_{r,r+2+k} = 0, k = 0, 1, 2, \cdots$, so that $b_r^\rho(\boldsymbol{\lambda}) = 0$. Repeating this argument for all ρ, $\Delta_1 U(\boldsymbol{\lambda}) = 0$ as claimed.[23]

Thus we are left with (4.23) with only the square bracket term on the rhs. Now we consider the $Z^{n3}(\boldsymbol{\lambda})$ poles of (4.4) which are singularities of $\mathcal{T}_3$ and $\overline{\mathcal{T}}_3$ only. Applying the decomposition procedure leading to (4.10), we find that $\Delta_3 \mathcal{T}(\boldsymbol{\lambda})(\mathbf{x}) = \Delta_3 \mathcal{T}(\boldsymbol{\lambda})(\mathbf{x})$. This is because, as noted earlier, $\overline{\mathcal{T}}^{\boldsymbol{\nu}}(\boldsymbol{\lambda})$ is obtained by complex conjugation with $\boldsymbol{\lambda}$ real, and then by analytic continuation; we have that $\Delta_3 \mathcal{T}(\boldsymbol{\lambda})(\mathbf{x})$ is concentrated on $x_1 = x_2 = x_3$ by Appendix B. Hence we get

$$0 = \mathcal{T}_3(\boldsymbol{\lambda})_R - \overline{\mathcal{T}}_3(\boldsymbol{\lambda})_R + \sum_{\substack{U \subset G \\ U \neq G, \neq \varphi}} (-1)^{|U|}\, \overline{\mathcal{T}}_{G(U)}(\boldsymbol{\lambda})_R \times \sum_{\substack{\nu_l = 1 \\ l \in M}}^{\infty} \left(\prod_{l \in M(U,U')} \frac{(g^2)^{(1-\lambda_l)\nu_l} \Delta_+^{\nu_l}(\lambda_l)}{\Gamma(\nu_l + 1)} \right) \times \mathcal{T}_{G(U')}(\boldsymbol{\lambda})_R . \tag{4.26}$$

Setting all the λ's equal to a single λ, each term admits analytic continuation to $\lambda = 0$. Performing the continuation, we derive the unitarity relation (4.13).[24]

The proof of the causality relation (4.14) follows on exactly the same lines. This completes the proof that the extension (4.11) and (4.12) is a renormalization.

5. THE UNIQUE EXTENSION WITH MINIMUM LIGHT CONE SINGULARITY

In this section we will show how a finite renormalization may be implemented so as to secure for $\mathcal{T}^{(3)}(\mathbf{x})$ an unique extension characterized by minimum light cone singularity.

(A) First we recall that $\mathcal{T}^{(2)}(\mathbf{x})$ has been completely fixed by virtue of unitarity, causality, and minimum light cone singularity. Then by virtue of the unitarity causality restrictions (4.13) and (4.14), which our extension for $\mathcal{T}^{(3)}(\mathbf{x})$ has been proved to satisfy, we have only the freedom to add to (4.11) a real, translation invariant distribution $\chi(\mathbf{x})$ from $\mathcal{C}_g'(R^{4\times 3})$ concentrated on $x_1 = x_2 = x_3$. In this subsection we will implement a specific finite renormalization. To this end we consider, in lieu of (4.11) and (4.12), the following:

Definition 5.1:

$$\mathcal{T}^{(3)}(x) = \mathfrak{F}^{-1}(\lim_{\eta \downarrow 0} \tilde{\mathcal{T}}_\eta^{(3)})(x), \tag{5.1}$$

where

$$\tilde{\mathcal{T}}_\eta^{(3)}(\mathbf{p}) = (8\pi)^2 \delta\left(\sum_1^3 P_i\right) \times \left\{ \left[\sum_{\nu_2, \nu_3 = 1}^{\infty} \prod_2^3 \left(\frac{(-1)^{\nu_l} (g^2)^{\nu_l}}{[(4\pi)^2]^{\nu_l} \Gamma(\nu_l) \Gamma(\nu_l + 1)} \right) \times \frac{1}{2i} \int_{\overline{\Gamma}_0} dz \left(\frac{g^2}{(4\pi)^2} \right)^z \frac{\cot \pi z (1 + \sin^2 \pi z)}{\Gamma(z) \Gamma(z+1)} \times e^{i\pi z} \frac{\Gamma(4 - z - \Sigma_2^3 \nu_l)}{2i \sin \pi(-z+1)} \times \int_0^1 dt_2\, t_2^{\nu_3 - 1} \int_1^{0+} dt_1 (-t_1)^{-z+1} [E(t_1, t_2)]^{2 - z - \Sigma_2^3 \nu_l} \times \sum_\rho [P_1^\rho(\mathbf{t}; p) + i\eta P_2^\rho(\mathbf{p})]^{z + \Sigma_2^3 \nu_l - 4} \right] + \tilde{\chi}(\mathbf{p}) \right\}, \tag{5.2}$$

where the contour $\overline{\Gamma}_0$ is shown in Fig. 9 and $\tilde{\chi}(p)$ is defined by[25]

$$\tilde{\chi}(p) \equiv 2 \left(\frac{g^2}{(4\pi)^2} \right)^2 \sum_\rho \int_0^1 dt_2 \int_0^1 dt_1 (1 + t_2 + t_2 t_1)^{-2} (t_1 t_2^{-1})^{1/2} \times \frac{1}{2\pi i} \int_C d\tau\, e^{t_1 \zeta_\rho \tau} I_1[2(1/\tau)^{1/2}] . \tau^2 \times I_1[2(t_2/\tau)^{1/2}] (1/\tau)^{1/2} K_1[2(1/t_1\tau)^{1/2}] \tag{5.3}$$

with $|\arg \tau| \leq \pi$, $\tau^{-1/2} = \exp(-\frac{1}{2} \log|\tau| - \frac{1}{2} i \arg \tau)$,

$$\zeta_\rho \equiv \frac{g^2}{(4\pi)^2} \left(\frac{p_{\rho(3)}^2 + t_1 t_2 p_{\rho(2)}^2 + t_1 p_{\rho(1)}^2}{t_1(1 + t_2 + t_2 t_1)} \right) \tag{5.4}$$

with standard Bessel functions[26] in (5.3).

The contour C is shown in Fig. 10. The contour is constrained, in the neighborhood of the origin, to be a cardioid; otherwise the τ integral, with the τ integrand analytic in the cut τ plane with the cut running along the real axis from 0 to $(-\infty)$, is contour independent.

With $\tau = R \exp i\theta$, we choose

$$R = \tfrac{1}{2} a (1 + \cos\theta) \tag{5.5}$$

for the contour C in the neighborhood of the origin, with any real $a > 0$. Then the real part of the arguments of the I_1 functions in (5.3) remains bounded in the neighborhood of the origin. Furthermore, since $|\arg \tau| \leq \pi$, it follows, from asymptotic estimates, that, for $0 \leq t_1 \leq 1$, $\tau^{-1/2} K_1[2(t_1\tau)^{-1/2}]$ remains bounded for $\tau \in C$. Thus the τ integral converges uniformly for $0 \leq t_{1,2} \leq 1$, and ζ_ρ in any compact set in $\mathbb{C}^1$. Since the τ integrand in (5.3) is analytic in the cut τ plane and (5.3) converges for all contour distortions with $|\arg| < \pi$ avoiding the neighborhood of the origin, by Cauchy's theorem (5.3) is

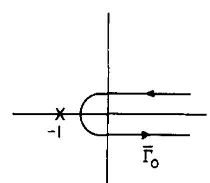

FIG. 9. The contour $\overline{\Gamma}_0$.

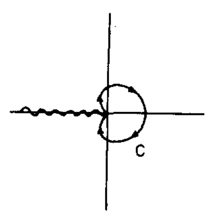

FIG. 10. The contour C.

contour independent. Hence (5.3) defines $\tilde{\chi}(p)$ unambiguously.

The definition (5.1) and (5.2) differs from the corresponding expressions (4.11) and (4.12) through the addition of real, translation invariant quasilocal distributions concentrated on $x_1 = x_2 = x_3$ which is an allowed finite renormalization. To verify this we first consider the term in braces in (5.2). This differs from the corresponding expression (4.12) in that (i) the contour $\overline{\Gamma}_0$ passes between zero and minus one, in contrast to $\overline{\Gamma}$ which passes between one and zero; (ii) there is the presence of an extra contribution from the $\sin^2 \pi z$ term[27] in the integrand. The latter contribution is due to *simple* z plane poles at the positive integers; it is easily evaluated and leads to an entire function of the (p_j^2) of order of growth $\frac{1}{3}$. Furthermore the difference in the contribution from the $\overline{\Gamma}_0$ and $\overline{\Gamma}$ contour integrals is again due to a *simple* pole at $z = 0$; the contribution is again an entire function of the p_j^2 of order of growth $\frac{1}{3}$.

It remains to characterize the contribution of $\tilde{\chi}(\mathbf{p})$ in (5.2). To this end we study

$$\mathcal{G}(\zeta_\rho, \mathbf{t}) = \frac{1}{2\pi i} \int_C d\tau e^{t_1 \zeta_\rho \tau} I_1 \left[2\left(\frac{1}{\tau}\right)^{1/2}\right] I_1\left[2\left(\frac{t_2}{\tau}\right)^{1/2}\right] \cdot \tau^2 \times \left(\frac{1}{t_1\tau}\right)^{1/2} K_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right], \tag{5.6}$$

$|\arg| \leq \pi$, which is the τ-contour integral in (5.3). The integrand is analytic in ζ_ρ; since we have a convergent integral with compact domain of integration, $(\zeta_\rho, \mathbf{t})$ is an entire function of ζ_ρ. We now estimate the order of growth in ζ_ρ.

For convenience let the entire contour C be the cardioid (5.5). Then for $\tau \in C$, we have $\mathrm{Re}(\tau^{-1/2}) = a^{-1/2}$. Also let $|\zeta_\rho| \leq r$. Then we have

$$\sup_{\zeta_\rho = r} |\mathcal{G}(\zeta_\rho)| \leq C_0 \cdot a \cdot e^{ra} \sup_{\tau \in C} \left\{ \left| I_1\left(2\left(\frac{1}{\tau}\right)^{1/2}\right) I_1\left(2\left(\frac{t_2}{\tau}\right)^{1/2}\right) \times \left(\frac{1}{\tau}\right)^{1/2} K_1\left(2\left(\frac{1}{t_1\tau}\right)^{1/2}\right)\right| \right\},$$

where C_0 is a constant independent of a, r. Let $a = f(r)$ (which will be determined) so that $a \to 0$, as $r \to \infty$, and hence $\mathrm{Re}(\tau^{-1/2}) \to \infty$. Then, for large r,

$$e^{ra} \sup_{\tau \in C} \{|\quad|\} \sim e^{ra} e^{2a^{-1/2}}$$

on using asymptotic estimates of Bessel functions.[26] We determine $a = f(r)$, by minimizing $g(a) = ra + 2a^{-1/2}$. $g'(a) = 0$ implies $a = r^{-2/3}$ and $g(a) = 3r^{1/3}$.

Hence

$$\lim_{r \to \infty} \exp(-3r^{1/3}) \sup_{\zeta_\rho = r} |\mathcal{G}(\zeta_\rho)| = 0. \tag{5.7}$$

Hence $\mathcal{G}(\zeta_\rho)$ is an entire function of ζ_ρ of order not greater than $\frac{1}{3}$. Recalling the definition (5.4) of ζ_ρ, it follows, because of the compact region of integration in t_1, t_2 in (5.3), that $\tilde{\chi}(\mathbf{p})$ is also an entire function of the (p_j^2) of order not greater than $\frac{1}{3}$.

We conclude that the difference between $\mathcal{T}^{(3)}(x)$, as defined in (5.1), and the corresponding object (4.11), which was obtained as a renormalization, is a real, translation invariant distribution in $\mathcal{C}_g'(R^{4\times 3})$ concentrated on $x_1 = x_2 = x_3$, which is an allowed finite renormalization.

(B) We shall now prove that the definition (5.1) of $\mathcal{T}^{(3)}(\mathbf{x})$ is unique and is characterized by minimum light cone singularity. We need only study $\mathrm{Re}\mathcal{T}^{(3)}(\mathbf{x})$, since $\mathrm{Im}\mathcal{T}^{(3)}(\mathbf{x})$ is fixed, by unitarity, any ambiguity being in the form of a real quasilocal distribution in $\mathcal{C}_g'(R^{4\times 3})$ concentrated on $x_1 = x_2 = x_3$ by virtue of causality. Returning to (5.2) we write

$$\mathcal{T}_\eta^{(3)}(p) = (8\pi)^2 \delta\left(\sum_{i=1}^3 p_i\right) \tilde{T}_\eta(\mathbf{p}) \tag{5.8}$$

and study $\tilde{T}_\eta(p)$ with the $\{p_j\}$ restricted to the region

$$\Omega = \{p_j | p_j^2 > 0,\ j = 1, 2, 3\}. \tag{5.9}$$

In compact subsets of this region $\lim_{\eta \downarrow 0} T_\eta(p)$ exists in the ordinary sense. This follows on recalling the definition in (3.20) of $P_1^\rho(\mathbf{t}, \mathbf{p})$ and on deforming the t_1 contour in (5.2) for each ρ term as in Fig. 11 with δ_ρ, the radius of the circular part satisfying

$$\delta_\rho < \left(\frac{p_{\rho(3)}^2}{p_{\rho(1)}^2 + p_{\rho(2)}^2}\right) \tag{5.10}$$

with the momenta restricted in any compact subset of Ω, (5.10) implying that $\mathrm{Re}P_1^\rho(\mathbf{t}; \mathbf{p}) > 0$. Hence we get[28] in any compact subset of Ω,

$$\mathrm{Re}[\lim_{\eta \downarrow 0} \tilde{T}_\eta(p)] = \left[\sum_{\nu_2, \nu_3 = 1}^{\infty} \prod_{l=2}^{3} \left(\frac{(-1)^{\nu_l}(g^2)^{\nu_l}}{[(4\pi)^2]^{\nu_l}\Gamma(\nu_l)\Gamma(\nu_l + 1)}\right) \times \frac{1}{2i}\int_{\overline{\Gamma}_0} dz \frac{[g^2/(4\pi)^2]^z}{\sin\pi z\, \Gamma(z)\Gamma(z+1)} \frac{\Gamma(4 - z - \Sigma_2^3 \nu_l)}{2i \sin\pi(-z+1)} \times \int_0^1 dt_2\, t_2^{\nu_3 - 1} \int_1^{0+} dt_1 (-t_1)^{-z+1} [E(t_1 t_2)]^{2 - z - \Sigma_2^3 \nu_l} \times \sum_\rho [P_1^\rho(\mathbf{t}; \mathbf{p})]^{z + \Sigma_2^3 \nu_l - 4}\right] + \tilde{\chi}(\mathbf{p}). \tag{5.11}$$

We can now deform the contour $\overline{\Gamma}_0 \to L$ (Fig. 12), the semicircular contributions vanishing at infinity, and the resulting integral converging along the imaginary axis. Furthermore, after the deformation $\overline{\Gamma}_0 \to L$, the t_1 Eulerian integral recovers its standard form, the circular contribution (Fig. 11) shrinking to zero as $\delta_\rho \to 0$. Thus

$$\mathrm{Re}[\lim_{\eta \downarrow 0} \tilde{T}_\eta(p)] = \left[\sum_{\nu_2, \nu_3 = 1}^{\infty} \prod_{l=2}^{3} \left(\frac{(-1)^{\nu_l} g^2/(4\pi)^{2\nu_l}}{\Gamma(\nu_l)\Gamma(\nu_l + 1)}\right) \times \frac{1}{2i}\int_L dz \frac{g^2/(4\pi)^{2z}}{\sin\pi z\, \Gamma(z)\Gamma(z+1)} \Gamma\left(4 - \sum_2^3 \nu_l - z\right) \times \int_0^1 dt_2 t_2^{\nu_3 - 1} \int_0^1 dt_1 t_1^{-z+1} [E(t_1 t_2)]^{2 - z - \Sigma_2^3 \nu_l} \times \sum_\rho [P_1^\rho(\mathbf{t}, \mathbf{p})]^{z + \Sigma_2^3 \nu_l - 4}\right] + \tilde{\chi}(\mathbf{p}). \tag{5.12}$$

In Appendix C we show that the double series within braces in (5.12) can be explicitly summed and prove that the lhs of (5.12) is given by

$$\mathrm{Re}[\tilde{T}(p)] = 2\left(\frac{g^2}{(4\pi)^2}\right)^2 \times \sum_\rho \int_0^1 dt_2 \int_0^1 dt_1\, (1 + t_2 + t_2 t_1)^{-2} (t_1 t_2^{-1})^{1/2}$$

$$\times \int_0^\infty d\tau\, e^{-t_1\zeta_\rho\tau} J_1\,[2(1/\tau)^{1/2}]\tau^2$$
$$\times\ J_1[2(t_2/\tau)^{1/2}](1/\tau)^{1/2} J_1[2(1/t_1\tau)^{1/2}] \qquad (5.13)$$

with ζ_ρ as defined in (5.4).

On changing variables $\tau \to (1/t_1)\tau$, we get

$$\mathrm{Re}(T(p)] = 2\left(\frac{g^2}{(4\pi)^2}\right)^2$$
$$\times \sum_\rho \int_0^1 dt_2 \int_0^1 dt_1\ (1+t_2+t_2t_1)^{-2} t_2^{-1/2}$$
$$\times \int_0^\infty d\tau\, e^{-\zeta_\rho\tau} J_1[2(t_1/\tau)^{1/2}]\,\tau^2 t_1^{-1}$$
$$\times J_1[2(t_1t_2/\tau)^{1/2}](1/\tau)^{1/2} J_1[2(1/\tau)^{1/2}]. \qquad (5.14)$$

We now consider the behavior of $\mathrm{Re}[\tilde{T}(p)]$ when the $p_j^2 \to +\infty$ specifically for two cases:

(i) $p_j^2 \to +\infty,\ j=1,2,3;$

(ii) $p_j^2, p_k^2 \to +\infty,\ p_l^2 \geq 0.$

In *either* case it follows from the definition (5.4) that $\zeta_\rho \to +\infty$, for all permutations ρ, *uniformly* in t_1, t_2 with $0 \leq t_{1,2} \leq 1$. Now the τ integral in (5.14) decreases faster than any inverse power of ζ_ρ, as $\zeta_\rho \to +\infty$. When $\zeta_\rho \to +\infty$, the dominant contribution is expected in the region $\tau \sim 0$; but in this region rapid oscillation sets in, independent of t_1, t_2, because of the last Bessel function in (5.14). An asymptotic estimate[29] shows that the actual decrease is $\exp(-\epsilon\zeta_\rho^{1/3})$, for some $\epsilon > 0$ independent of the t_j. Since the region of $t_{1,2}$ integrations is compact, and since the integrand (5.14) is strongly decreasing in cases (i) and (ii) above for all $t_i, 0 \leq t_i \leq 1$, *it follows that Re T(p) is a strongly decreasing function in both of the above cases (i) and (ii)*. We are now in a position to prove the following:

Theorem 5.1: The extension $\mathcal{T}^{(3)}(x)$, as given in Definition 5.1, is unique and has minimum light cone singularity.

The proof of uniqueness follows immediately from the previously established strong decrease in the case (i) $p_j^2 \to +\infty$, $j = 1, 2, 3$. For, as has been mentioned earlier, the only arbitrariness is in $\mathrm{Re}\mathcal{T}_3(x)$ to which we can add a real, translation invariant distribution from $\mathcal{C}_g'(R^{4\times 3})$ concentrated on $x_1 = x_2 = x_3$. In momentum space, after factoring the over-all function of p_j^2 of order $< \frac{1}{2}$ which cannot decrease to zero in any direction and would destroy the asymptotic property in the region (i).

From the (stronger) property of decrease in the region, (ii)[30] follows the absence in $\mathrm{Re}\mathcal{T}_3(\mathbf{x})$ of distributions concentrated on $x_1 = x_2 = x_3$. To see this, define

$$\mathrm{Re}\tilde{\mathcal{T}}_3(p) \equiv \delta^{(4)}(p_1+p_2+p_3)\tilde{T}(p_1^2, p_2^2, p_3^2), \quad (5.15)$$

so that

$$\mathrm{Re}\mathcal{T}_3(x_1x_2x_3) = T(\xi, \eta), \qquad (5.16)$$

where

$$T(\xi,\eta) = \int_{R^4} dp_2 \int_{R^4} dp_3\ e^{-i(p_2\xi+p_3\eta)} T((p_2+p_3)^2, p_2^2, p_3^2)$$

and

$$\xi = x_1 - x_2,\ \eta = x_1 - x_3.$$

Let $\xi_\pm^0 = \frac{1}{2}(\xi^0 \pm \eta^0)$ and define

$$t(\xi_-^0) \equiv \int_{R^7} d\vec{\xi}\, d\vec{\eta}\, d\xi_+^0\ T(\xi,\eta)\phi(\xi,\eta,\xi_+^0) \qquad (5.17)$$

and where $\phi \in \mathcal{C}_g(R^7)$ and $\tilde{\phi}(\vec{p}_2, \vec{p}_3, p_+^0) \in \mathfrak{M}_g(R^7)$ and of compact support,

$$\mathrm{Supp}\ \tilde{\phi} = \left\{\vec{p}_2, \vec{p}_3, p_+^0 \,\middle|\, \begin{matrix} \|\vec{p}_{2,3}\| \leq L \\ 2L \leq |p_+^0| \leq M \end{matrix}\right\}, \qquad (5.18)$$

where $p_\pm^0 = p_2^0 \pm p_3^0$.

Then we have the identity

$$t(\xi_-^0) = \int_{R^1} dp_-^0\, e^{-ip_-^0\xi^0}\tilde{t}(p_-^0), \qquad (5.19)$$

where

$$\tilde{t}(p_-^0) = \int_{R^3} d\vec{p}_2 \int_{R^3} d\vec{p}_3 \int_{R^1} dp_+^0 \tilde{T}((p_2+p_3)^2, p_2^2, p_3^2)$$
$$\tilde{\phi}(\vec{p}_2, \vec{p}_3, p_+^0). \qquad (5.20)$$

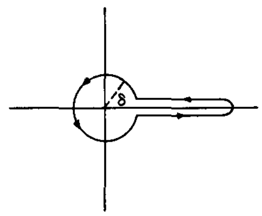

FIG. 11. The t_1 contour.

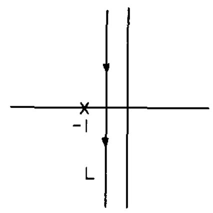

FIG. 12. The L contour.

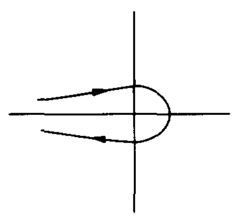

FIG. 13. The Hankel contour.

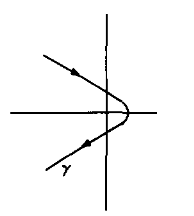

FIG. 14. The contour γ.

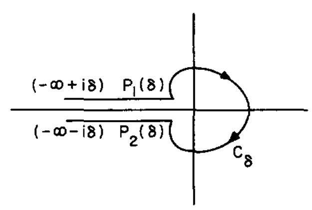

FIG. 15. The contour C_δ.

Writing

$$\left.\begin{aligned}(p_2+p_3)^2 &= p_+^{0\,2} - (\vec{p}_2+\vec{p}_3)^2\\ p_2^2 &= \tfrac{1}{4}(p_+^0+p_-^0)^2 - \vec{p}_2^{\,2}\\ p_3^2 &= \tfrac{1}{4}(p_+^0-p_-^0)^2 - \vec{p}_3^{\,2}\end{aligned}\right\}, \quad \text{let } |p_-^0| \to \infty.$$

Then we get, because of (5.18), $\zeta_\rho \to +\infty$ in the contributing region in (5.20) because of case (ii) after (5.14). Since $\tilde{T}$ is strongly decreasing, because of compact region of integration in (5.20) so is $\tilde{t}(p_-^0)$. Hence $t(\xi_-^0)$ is C^∞ at $\xi_-^0 = 0$. This rules out the presence of distributions like $P(\partial/\partial\xi, \partial/\partial\eta)\delta^{(4)}(\xi)\delta^{(4)}(\eta)$ [but not like $P(\partial/\partial\xi)\delta^{(4)}(\xi)$ or $P(\partial/\partial\eta)\delta^{(4)}(\eta)$ which is allowed]. Hence the unique renormalized three-point function (5.1), (5.2) is characterized by minimum lightcone singularity. QED

6. CONCLUDING REMARKS

In this paper we have taken the point of view that (axiomatic) renormalization[2,3] is basic in local field theory. We showed how some time-ordered products in a conventionally nonrenormalizable but strictly local field theory can be uniquely defined by first effecting a renormalization and then by imposing, consistently, a light cone boundary condition to fix the allowed finite renormalization. Divergences arise in naive manipulations in field theory due to locality; nonformal methods lead to a variety of extensions. Therefore it seems to us highly sensible to impose, if possible, a boundary condition on the light cone to fix the "simplest" dynamics[5] and to secure a unique extension with the minimum allowed light cone singularity. It is due to remarkable properties of infinite sets of renormalized Feynman graphs that this situation obtains for the two- and three-point T products of this paper; it is characteristic of the "disappearance" of the effects of logarithms (in the kinematic invariants) in the sums. Our results encourage us to believe that higher point functions can also be uniquely renormalized with minimum light cone singularity, although complete results in this direction are not yet available. A successful implementation of this program would lead to a control, in perturbation theory, of at least a privileged class of nonrenormalizable fields in accord with the general principles of local field theory.

Note added in proof: K. Pohlmeyer[13] proves minimum singularity with pure space smearing. The equality of our definition (5.1–5.2) with Pohlmeyer's follows on inserting Mellin–Barnes representations for Bessel functions in (5.14) and performing the τ integral. We thank R. Flume and K. Pohlmeyer for communications.

ACKNOWLEDGMENTS

P.K.M. thanks Professor H. Lehmann for his continued interest and for valuable conversations and correspondence during the course of this investigation. He also thanks Professor D.I. Fivel for stimulating conversations and Professor R.E. Peierls for his hospitality at Oxford where most of the work was done.

APPENDIX A

In this appendix we verify some assertions of Sec. 3D.

(I) To begin with take any $\eta > 0$, and consider the representation (3.28) for $A(\nu/\eta)$. Suppressing the dependence on $\boldsymbol{\lambda}$ and t_2, we have

$$A_\eta^{\boldsymbol{\nu}} = \int_C dt_1 G^{(\nu)}(t_1)[P_1(\mathbf{p}, t_1) + i\eta P_2(\mathbf{p})]^{-(\mu_3+4)},$$

where the $\delta(\sum p)$ restriction is understood. $G^{\boldsymbol{\nu}}(\mathbf{t})$ is analytic on C with growth (in $\boldsymbol{\nu}$) bounded by $c^{\|\nu\|}$, P_2 is positive definite, and it is permissible to write

$$\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle = \int_C dt_1\, G^{(\nu)}(t_1)\langle (P_1(t_1) + i\eta P_2)^{-(\mu_3+4)}, \tilde{\phi}\rangle,$$

with $\phi \in \mathfrak{M}_g(R^{4\times 2})$. We have

$$\begin{aligned}|\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| &\le L_\eta \sup_{t_1\in C} G^{\boldsymbol{\nu}}(t_1)|\sup_{t_1\in C}| \\ &\qquad \times \langle [P_1(t_1) + i\eta P_2]^{-(\mu_3+4)}, \tilde{\phi}\rangle| \\ &\le L_\eta C_1^{\|\nu\|}\int_{R^{4\times 2}} dp\,|\tilde{\phi}(p)|\sup_{t_1\in C}| \\ &\qquad \times [P_1(p, t_1) + i\eta P_2(p)]^{a\|\nu\|_3},\end{aligned}$$

with $0 < \mathrm{Re}\,a < 1$, and for large $\|\nu\|_3$ can ignore -4, which can always be compensated by integration by parts (Sec. 3C). Thus

$$\begin{aligned}|\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| &\le L_\eta C_2^{\|\nu\|}\|\tilde{\phi}\|_g \int_{R^{4\times 2}} dp\; g(\|p\|^2)^{-1} \\ &\qquad (\sup_{t_1\in C}|P_1(p, t_1)| + \eta P_2(p))^{\|\nu\|_3},\end{aligned}$$

where $g(\|p^2\|)$ is the indicator function[1] and $\|\tilde{\phi}\|_g$ is a norm in $_g(R^{4\times 2})$. Introducing hyperspherical coordinates in $R^{4\times 2}$, we get

$$\tilde{\psi}_\eta^{(\boldsymbol{\nu})} = C_2^{\|\nu\|}\int_{\text{compact region}} d\Omega[\hat{P}_1(\Omega) + \hat{P}_2(\Omega)]^{\|\nu\|_3}$$

with

$$|\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| \le L_\eta \|\tilde{\phi}\|_g \tilde{\psi}^{(\boldsymbol{\nu})}\int_0^\infty dR\, R^7[g(R^2)]^{-1}(R^2)^{\|\nu\|_3}$$

whence, utilizing the decrease property[1] of $[g(R^2)]^{-1}$, we get

$$|\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| \le L_\eta \tilde{\psi}_\eta^{(\boldsymbol{\nu})}\|\nu\|_3^{[(1/\rho)\|\nu\|_3]}\|\tilde{\phi}\|_g, \tag{A1}$$

where ρ is the order of growth of $g(R^2)$. Obviously $\psi_\eta^{(\boldsymbol{\nu})} \sim (\text{const})^{\|\nu\|_3}$ for large $\|\nu\|_3$.

(II) We now show that the bound continues to hold in the limit $\eta \downarrow 0$. As in Sec. 3C, we return to the equivalent representation (3.24). Then

$$\begin{aligned}A_\eta^{\boldsymbol{\nu}} = \Bigg[&\sum_{k_1=0}^{\nu_1}\left(\frac{(-1)^{k_1}(\partial/\partial t_1)^{k_1}\big|_{t_1=1}}{\Pi_{m_1=0}^{k_1}(\mu_1+2+m_1)}\right) \\ &+ \frac{(-1)^{\nu_1}}{\Pi_{m_1=0}^{\nu_1}(\mu_1+2+m_1)}\int_0^1 dt_1\, t_1^{\mu_1+2+\nu_1}\left(\frac{\partial}{\partial t_1}\right)^{\nu_1+1}\Bigg] \\ &\times \{E_\rho(\mathbf{t})^{\mu_3+2}[P_1(\mathbf{p}, t) + i\eta P_2(\mathbf{p})]^{-(\mu_3+4)}\},\end{aligned}$$

which consists of $\nu_1 + 1$ terms. Because of the action of derivatives, each term is a sum of at most $\nu_1 + 1$ terms of the form (assume large $\|\nu\|$)

$$T_\eta^{\boldsymbol{\nu}} \sim (\text{const})^{\nu_1}F_\omega(p^2, t)[P_1(p, t) + i\eta P_2(p)]^{a\|\nu\|_3-\omega-4},$$

where $F_w(p^2)$ is a polynomial in the $p_i\cdot p_j$ of degree $w \le \nu_1 + 1$, and $0 < \mathrm{Re}\,a < 1$. In smearing A_η^{ν} with a test function $\tilde{\phi} \in \mathfrak{M}_g(R^{4\times 2})$, we have to consider (with large $\|\nu\|$)

$$\langle (P_1 + i\eta P_2)^{a\|\nu\|_3 - w - 4}, F_w \tilde{\phi}\rangle \sim \left(\frac{\Gamma(a\|\nu\|_3)}{\Gamma(a\|\nu\|_3 - w)}\right)^{-2} \times \langle (P_1 + i\eta P_2)^{a\|\nu\|_3}, \mathfrak{L}^{w+4}(F_w\tilde{\phi})\rangle$$

on utilizing (3.17). The right-hand side is continuous in η. One readily obtains

$$|\langle T_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| \le C\|\tilde{\phi}\|_g \tilde{\psi}_\eta^{(\nu)} \|\nu\|_3^{(1/\rho)(\|\nu\|_3 + \rho w)} \|\nu\|_3^{-2w} \le C\|\tilde{\phi}\|_g \tilde{\psi}_\eta^{(\nu)} \|\nu\|_3^{(1/\rho)\|\nu\|_3}$$

for large $\|\nu\|_3$ where $\tilde{\psi}_\eta^{(\nu)}$ is continuous in η and so we obtain easily the bound

$$|\langle A_\eta^{\boldsymbol{\nu}}, \tilde{\phi}\rangle| \le C'\|\tilde{\phi}\|_g \tilde{\psi}_{2(\eta)}^{(\nu)} \|\nu\|_3^{(1/\rho)\|\nu\|_3} \tag{A2}$$

for large $\|\nu\|_3$ with $\tilde{\psi}_{2(\eta)}^{(\nu)} \sim (\mathrm{const})^{\|\nu\|_3}$ and continuous in η in the neighborhood of zero. This suffices to establish the uniformity of convergence stated after (3.28).

APPENDIX B

In this Appendix we verify the support properties of the singular (in $\boldsymbol{\lambda}$) pieces $\Delta_j \mathcal{T}_\eta(\boldsymbol{\lambda})$ of (4.10) stated after that equation.

First consider $\Delta_3\mathcal{T}_\eta(\boldsymbol{\lambda})$. We have, by definition,

$$\Delta_3\mathcal{T}_\eta(\boldsymbol{\lambda}) = (8\pi)^2\delta\left(\sum_1^3 p_i\right)\sum_{\nu_2,\nu_3=1}^{\infty}\frac{1}{\Gamma(\nu_2+1)\Gamma(\nu_3+1)} \times 2\pi i\sum_\rho \sum_{n_3 \ge \max\{0,\nu_2+\nu_3-4\}} \operatorname*{res}_{z=z^{n_3}}[z \text{ integrand of } (4.2)], \tag{B1}$$

where z^{n_3} is given in (4.4). Hence

$$\Delta_{3\,\eta\downarrow 0}(\boldsymbol{\lambda}) = (8\pi)^2\delta\left(\sum_1^3\ \right)\sum_{\nu_2,\nu_3}\frac{1}{\Gamma(\nu_2+1)\Gamma(\nu_3+1)} \times \sum_\rho \prod_{l=2}^3 \frac{2^{2\lambda_{\rho(l)}\nu_l}e^{(\lambda_{\rho(l)}-1)\nu_l\pi i}}{\Gamma((1-\lambda_{\rho(l)})\nu_l)} \sum_{n_3\ge\max\{0,\nu_2+\nu_3-4\}}\frac{\pi(-1)^{n_3}}{\Gamma(n_3+1)}$$
$$\times \frac{\cot\pi z^{n_3}}{\Gamma(z^{n_3})\Gamma(z^{n_3}+1)}\frac{2^{2\lambda_{\rho(1)}z^{n_3}}e^{(\lambda_{\rho(1)}-1)z^{n_3}\pi i}}{[(4\pi)^2]^{z^{n_3}}}(g^2)^{n_3} \times \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_3]-1}\{2i\sin\pi[(\lambda_{\rho(1)}-1)\nu_1+1]\}^{-1}$$
$$\times \int_1^{0+} dt_1(-t_1)^{(\lambda_{\rho(1)}-1)z^{n_3}}[E(t_1t_2)]^{-(n_3+2)}[P_1^\rho(\mathbf{p},t)]^{n_3}. \tag{B2}$$

Omitting the $\delta(\sum_1^3 p_i)$, (B2) represents an entire function of all the p_j^2 of order $<\frac{1}{2}$. Hence $\mathfrak{F}^{-1}[\Delta_3\tilde{\mathcal{T}}_{\eta\downarrow 0}(\boldsymbol{\lambda})](\mathbf{x})$ is a distribution in $\mathcal{C}_g'(R^{4\times 3})$ concentrated on $x_1 = x_2 = x_3$.

Next we consider $\Delta_1\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda})$. According to its definition

$$\Delta_1\tilde{\mathcal{T}}_\eta(\boldsymbol{\lambda}) = (8\pi)^2\delta\left(\sum_1^3 p_i\right)\sum_{\nu_2,\nu_3=1}^{\infty}\frac{1}{\Gamma(\nu_2+1)\Gamma(\nu_3+1)} \times 2\pi i\sum_\rho\sum_{n_1=0}^{\infty}\operatorname*{res}_{z=z^{n_1}}[z \text{ integrand of } (4.2)], \tag{B3}$$

where z^{n_1} is defined in (4.4). Hence

$$\Delta_1\tilde{\mathcal{T}}_{\eta\downarrow 0}(\boldsymbol{\lambda}) = (8\pi)^2\delta\left(\sum_1^3 p_i\right)\sum_\rho\sum_{n_1=2}^{\infty} a_{n_1}^\rho(\lambda_{\rho(1)}) \times \mathcal{P}_{n_1}^\rho(\lambda_{\rho(2)},\lambda_{\rho(3)})(\mathbf{p}) \tag{B4}$$

where

(i) $$a_{n_1}^\rho(\lambda_{\rho(1)}) = \frac{2(2\lambda_{\rho(1)}n_1)/(1-\lambda_{\rho(1)})(-1)^{n_1}}{[(4\pi)^2]^{(n_1)/(1-\lambda_{\rho(1)})}\Gamma(n_1)\Gamma\{[n_1/(1-\lambda_{\rho(1)})]+1\}} \times \cot\frac{\pi n_1}{1-\lambda_{\rho(1)}},$$

(ii) $$\mathcal{P}_{n_1}^\rho(\lambda_{\rho(2)},\lambda_{\rho(3)})(\mathbf{p}) = \sum_{\nu_2,\nu_3=1}^{\infty}\prod_{l=2}^3\left(\frac{2^{\lambda_{\rho(l)}\nu_l}e^{(\lambda_{\rho(l)}-1)\nu_l\pi i}(g^2)^{(1-\lambda_{\rho(l)})\nu_l}}{\Gamma(\nu_l+1)\Gamma((1-\lambda_{\rho(l)})\nu_l)}\right) \times (g^2)^{n_1}\Gamma\left(4-n_1-\sum_2^3(1-\lambda_{\rho(l)})\nu_l\right)$$
$$\times \int_0^1 dt_2\, t_2^{-[(\lambda_{\rho(3)}-1)\nu_3]-1}\frac{1}{2i}\lim_{\eta\to 0^+}\int_1^{0+}dt_1\, t_1^{-n_1+1} \times [E(t_1t_2)]^{2-n_1-\Sigma_2^3(1-\lambda_{\rho(l)})\nu_l} \times [P_1^\rho(\mathbf{p},t)+i\eta P_2^\rho(\mathbf{p})]^{\Sigma_2^3(1-\lambda_{\rho(l)})\nu_l+n_1-4}.$$

Now the t_1 contour integral can be evaluated since we have a pole of order (n_1-1), and all other factors are analytic; its evaluation gives

(iii) $$\frac{2\pi i}{(n_1-2)!}\sum_{r=0}^{n_1-2}\binom{n_1-2}{r}\left[\left(\frac{\partial}{\partial t_1}\right)^{n_1-2-r} \times [E(t_1t_2)]^{2-n_1-\Sigma_2^3(1-\lambda_{\rho(l)})\nu_l}\right] \times \left[\left(\frac{\partial}{\partial t_1}\right)^r[P_1^\rho(\mathbf{p},t)+i\eta P_2^\rho(\mathbf{p})]_{\eta\downarrow 0}^{[\Sigma_2^3(1-\lambda_{\rho(l)})\nu_l+n_1-4]}\right]$$

evaluated at $t_1 = 0$. Recalling that $P_1(\mathbf{p},t) = p_{\rho(3)}^2 + t_1t_2p_{\rho(2)}^2 + t_1p_{\rho(1)}^2$, the last factor in braces in (iii) gives a polynomial of degree "r" in $p_{\rho(2)}^2, p_{\rho(1)}^2$. Hence $\mathfrak{F}^{-1}[\delta(\sum p)\times \text{last }\{\ \}_{\eta\downarrow 0}]$ is concentrated on $x_{\rho(1)} = x_{\rho(2)}$. Putting (i), (ii), and (iii) in (B4), we have that $\mathfrak{F}^{-1}[\Delta_{1\,\eta\downarrow 0}(\boldsymbol{\lambda})](\mathbf{x})$ consists of a sum of terms, each one a convergent (in $\mathcal{C}_g'$) series of distributions concentrated on $x_{\rho(1)} = x_{\rho(2)}$, etc.

APPENDIX C

In this appendix[31] we prove (5.13) starting from (5.12). Defining $\lim_{\eta\downarrow 0} T_\eta(p) \equiv T(p)$, we write

$$\operatorname{Re}\tilde{T}(p) = \tilde{F}(p) + \tilde{\chi}(p), \tag{C1}$$

where $\tilde{F}(p)$ represents the term in braces in (5.12). We have

$$\tilde{F}(p) = (-\pi)\sum_{\nu_2,\nu_3=1}^{\infty}\sum_{l=2}^{3}\left(\frac{[g^2/(4\pi)^2]^{\nu_l}}{\Gamma(\nu_l)\Gamma(\nu_l+1)}\right) \times \frac{1}{2i}\int_L dz\,\frac{[g^2/(4\pi)^2]^z}{\sin^2\pi z\,\Gamma(z)\Gamma(z+1)\Gamma(z+\Sigma_2^3\nu_l-3)}$$
$$\times \int_0^1 dt_2\, t_2^{\nu_3-1}\int_0^1 dt_1\, t_1^{-z+1}(1+t_2+t_2t_1)^{2-z-\Sigma_2^3\nu_l} \times \sum_\rho[p_{\rho(3)}^2+t_1t_2p_{\rho(2)}^2+t_1p_{\rho(1)}^2]^{(z+\Sigma_2^3\nu_l-4)} \tag{C2}$$

with the contour L as given in Fig. 12. Interchanging the z integration with the t_1, t_2 integrations, which is valid due to uniform convergence, introducing the variable ζ_ρ defined in (5.4), and recalling that $\sum_\rho$ is a finite sum, we have

$$\tilde{F}(p) = (-\pi)\left(\frac{g^2}{(4\pi)^2}\right)^2 \sum_\rho \sum_{\nu_2,\nu_3=1}^{\infty} \prod_{l=2}^{3} \left(\frac{1}{\Gamma(\nu_l)\Gamma(\nu_l+1)}\right) \times \int_0^1 dt_2\, t_2^{\nu_3-1} \int_0^1 dt_1 (1+t_2+t_2 t_1)^{-2} t_1 (t_1\zeta_\rho)^{\nu_2+\nu_3-4} \times \frac{1}{2i}\int_L dz \frac{(\zeta_\rho)^z}{\sin^2\pi z\,\Gamma(z)\Gamma(z+1)\Gamma(z+\Sigma_2^3\nu_l-3)}. \quad \text{(C3)}$$

(The dependence of ζ_ρ on t_1, t_2, and $\{p_j^2\}$ should be kept in mind.) It is also permissible, due to uniform convergence, to interchange the t_1, t_2 integration with the double series summation, so that we have

$$\tilde{F}(p) = (-\pi)\left(\frac{g^2}{(4\pi)^2}\right)^2 \sum_\rho \int_0^1 dt_2 \int_0^1 dt_1 (1+t_2+t_2t_1)^{-2} \times t_1 (t_1\zeta_\rho)^{-4}\tilde{G}(\mathrm{t},\zeta_\rho), \quad \text{(C4)}$$

where

$$\tilde{G}(\mathrm{t},\zeta_\rho) \equiv \sum_{\nu_2,\nu_3=1}^{\infty} \prod_{l=2}^{3}\left(\frac{(t_1\zeta_\rho)^{\nu_l}}{\Gamma(\nu_l)\Gamma(\nu_l+1)}\right) t_2^{\nu_3-1}\frac{1}{2i}\int_L dz \times \frac{(\zeta_\rho)^z}{\sin^2\pi z\,\Gamma(z)\Gamma(z+1)\Gamma(z+\Sigma_2^3\nu_l-3)} \quad \text{(C5)}$$

$$= \frac{1}{2i}\int_L dz \frac{(\zeta_\rho)^z}{\sin^2\pi z\,\Gamma(z)\Gamma(z+1)}\left[\sum_{\nu_2,\nu_3=1} \times \prod_{l=2}^{3}\left(\frac{(t_1\zeta_\rho)^{\nu_l}}{\Gamma(\nu_l)\Gamma(\nu_l+1)}\right) t_2^{\nu_3-1}\frac{1}{\Gamma(z+\Sigma_2^3\nu_l-3)}\right], \quad \text{(C6)}$$

the last interchange being valid due to uniform convergence, as follows readily via Stirling's formula. We now introduce in (C6) the integral representation[15]

$$\frac{1}{\Gamma(z+\nu_2+\nu_3-3)} = \frac{1}{2\pi i}\int_{-\infty}^{0+} d\tau\, e^\tau \tau^{-(z+\nu_2+\nu_3-3)} \quad \text{(C7)}$$

with $|\arg\tau| \le \pi$, τ^{-z} being defined as $\exp[-z(\log|\tau| + i\arg\tau)]$. The contour in (C7) begins and ends at $(-\infty)$ clockwise encircling the origin once (see Fig. 13). It is convenient for the subsequent development to replace (C7) by

$$\frac{1}{\Gamma(z+\nu_2+\nu_3-3)} = \frac{1}{2\pi i}\int_\gamma d\tau\, e^\tau \tau^{-(z+\nu_2+\nu_3-3)} \quad \text{(C8)}$$

with the deformed contour γ submitting to $\frac{1}{2}\pi < |\arg\tau| < \pi$ as shown in Fig. 14, the deformation being valid due to strong decrease at infinity.

Introducing the representation (C8) in (C6) we interchange, by virtue of uniform convergence, the double series summation with the τ integration, to get

$$\tilde{G}(\mathrm{t},\zeta_\rho) = \frac{1}{2i}\int_L dz \frac{(\zeta_\rho)^z}{\sin^2\pi z\,\Gamma(z)\Gamma(z+1)} \times \frac{1}{2\pi i}\int_\gamma d\tau e^\tau \tau^{3-z}\left(\sum_{\nu_2=1}^{\infty}\frac{[(t_1\zeta_\rho)/\tau]^{\nu_2}}{\Gamma(\nu_2)\Gamma(\nu_2+1)}\right) \times \left(\frac{1}{t_2}\sum_{\nu_3=1}^{\infty}\frac{[(t_1t_2\zeta_\rho/\tau)]^{\nu_3}}{\Gamma(\nu_3)\Gamma(\nu_3+1)}\right)$$

$$= t_1 t_2^{-1/2}\zeta_\rho \frac{1}{2i}\int_L dz \frac{(\zeta_\rho)^z}{\sin^2\pi z\,\Gamma(z)\Gamma(z+1)} \times \frac{1}{2\pi i}\int_\gamma d\tau\, e^\tau \tau^{-z} I_1\left[2\left(\frac{t_1\zeta_\rho}{\tau}\right)^{1/2}\right] I_1\left[2\left(\frac{t_1t_2\zeta_\rho}{\tau}\right)\right]^{1/2}.\tau^2 \quad \text{(C9)}$$

on making appropriate indentifications of the two series in braces with Bessel functions.[26]

Finally, on noting the restriction $\frac{1}{2}\pi < |\arg\tau| < \pi$ for $\tau \in \gamma$, and that ζ_ρ is real and positive, the τ and z integrations in (C9) may be interchanged due to uniform convergence. In fact the convergence of the τ integral is independent of $z (z \in L)$ and the convergence of the z integral is uniform in τ because $|\arg\tau| < \pi$, $\tau \in \gamma$. By performing the interchange and recognizing[32] that

$$-\frac{1}{2\pi i}\int_L dz\Gamma(1-z)\Gamma(-z)\left(\frac{\zeta_\rho}{\tau}\right)^z = 2\left(\frac{\zeta_\rho}{\tau}\right)^{1/2} K_1\left[2\left(\frac{\zeta_\rho}{\tau}\right)^{1/2}\right],$$

where K_1 is the modified Bessel function, we get, on making the change of variable $\tau \to t_1\zeta_\rho\tau$,

$$\tilde{G}(\mathrm{t},\zeta_\rho) = \frac{1}{\pi}(t_1\zeta_\rho)^4 t_2^{-1/2}\frac{1}{2\pi i}\int_\gamma d\tau\, e^{t_1\zeta_\rho\tau} I_1\left[2\left(\frac{1}{\tau}\right)^{1/2}\right].\tau^2 \times I_1\left[2\left(\frac{t_2}{\tau}\right)^{1/2}\right] 2\left(\frac{1}{t_1\tau}\right)^{1/2} K_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right]. \quad \text{(C10)}$$

Next we deform the contour γ of Fig. 14 to that of Fig. 15.

The points $P_1(\delta), P_2(\delta)$ are the intersections of the two straight portions, parellel to the real axis, with the curve (5.5). We choose, for convenience, the portion C_δ of the contour, starting at $P_1(\delta)$ and ending at $P_2(\delta)$, to lie on (5.5). We express (C10) as the sum of two parts:

$$\tilde{G}(\mathrm{t},\zeta_\rho) = \tilde{G}_\delta^{(1)} + \tilde{G}_\delta^{(2)}, \quad \text{(C11)}$$

where $\tilde{G}_\delta^{(1)}$ receives its contribution from the portions $(-\infty + i\delta), P_1(\delta))$ and $(P_2(\delta), -\infty - i\delta)$, and $\tilde{G}_\delta^{(2)}$ receives its contribution from the portion C_δ of the contour.

We shall now take the limit $\delta \to 0$. In the limit, $P_{1,2}(\delta) \to 0$ and we recover $C_\delta \to C$, the contour of (5.5). Then

$$\lim_{\delta\to 0}\tilde{G}_\delta^{(2)} = \frac{2}{\pi}(t_1\zeta_\rho)^4 t_2^{1/2}\frac{1}{2\pi i}\int_C d\tau\, e^{t_1\zeta_\rho\tau} I_1\left[2\left(\frac{1}{\tau}\right)^{1/2}\right] \times I_1\, 2\left(\frac{t_2}{\tau}\right)^{1/2}\left(\frac{1}{t_1\tau}\right)^{1/2} K_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right]. \quad \text{(C12)}$$

The existence and contour independence (avoiding distortion in the neighborhood of the origin) follows from the arguments after (5.4). The limit as $\delta \to 0$ of $\tilde{G}_\delta^{(1)}$ also exists. To calculate it we need to compute the discontinuity of the integrand of (C10). Now the product of the two I_1 functions in (C10) is analytic in the punctured (at the origin) τ plane, whereas for the nonanalytic part we use the identity[33]

$$\frac{2K_1[2(1/t_1\tau)^{1/2}]}{2(1/t_1\tau)^{1/2}} = K_0\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right] + K_2\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right],$$

together with[26]

$$K_{2n}\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right]_{\text{nonanalytic piece}}$$

$$= - I_{2n}\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right] \log\tau^{-1/2}$$

for the piece contributing to the discontinuity to get

$$\operatorname{disc}\left(\frac{2K_1[2(1/t_1\tau)^{1/2}]}{2(1/t_1\tau)^{1/2}}\right) = -2\pi i\left(\frac{2I_1[2(1/t_1\tau)^{1/2}]}{2(1/t_1\tau)^{1/2}}\right)$$

on using the similar identity for the I_n functions. Thus we get

$$\lim_{\delta\to 0} \tilde{G}_\delta^{(1)} = \frac{1}{\pi}(t_1\zeta_\rho)^4 t_2^{1/2}\int_0^{-\infty} d\tau\, e^{t_1\zeta_\rho\tau} I_1\left[2\left(\frac{1}{\tau}\right)^{1/2}\right]$$
$$\times 2\left(\frac{1}{t_1\tau}\right)^{1/2} I_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right] I_1\left[2\left(\frac{t_2}{\tau}\right)^{1/2}\right]$$
$$= -\frac{2}{\pi}(t_1\zeta_\rho)^4 t_2^{1/2}\int_0^{\infty} d\tau\, e^{-t_1\zeta_\rho\tau} J_1\left[2\left(\frac{1}{\tau}\right)^{1/2}\right]$$
$$\times J_1\left[2\left(\frac{t_2}{\tau}\right)^{1/2}\right]\left(\frac{1}{t_1\tau}\right)^{1/2} J_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right], \qquad \text{(C13)}$$

which also exists. Hence from (C11)–(C13), we get

$$\tilde{G}(t,\zeta_\rho) = -\frac{2}{\pi}(t_1\zeta_\rho)^4 t_2^{1/2}\int_0^{\infty} d\tau\, e^{-t_1\zeta_\rho\tau} J_1\left[2\left(\frac{1}{\tau}\right)^{1/2}\right]$$
$$\times J_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right] J_1\left[2\left(\frac{t_2}{\tau}\right)^{1/2}\right]\left(\frac{1}{t_1\tau}\right)^{1/2}$$
$$+\frac{2}{\pi}(t_1\zeta_\rho)^4 t_2^{1/2}\frac{1}{2\pi i}\int_C d\tau\, e^{t_1\zeta_\rho\tau} I_1\left[2\left(\frac{1}{\tau}\right)^{1/2}\right]$$
$$\times I_1\left[2\left(\frac{t_2}{\tau}\right)^{1/2}\right]\left(\frac{1}{t_1\tau}\right)^{1/2} K_1\left[2\left(\frac{1}{t_1\tau}\right)^{1/2}\right]. \qquad \text{(C14)}$$

Combining (C14) with (C4), and noting the definition (5.3) of $\tilde{\chi}(p)$, we get

$$\tilde{F}(p) = 2\left(\frac{g^2}{(4\pi)^2}\right)^2$$
$$\times \sum_\rho \int_0^1 dt_2 \int_0^1 dt_1\, (1+t_2+t_2t_1)^{-2}(t_1t_2)^{1/2}$$
$$\times \int_0^\infty d\tau\, e^{-t_1\zeta_\rho\tau} J_1[2(1/\tau)^{1/2}] J_1[2(t_2/\tau)^{1/2}](1/\tau)^{1/2}$$
$$\times J_1[2(1/t_1\tau)^{1/2}] - \tilde{\chi}(\mathrm{p}). \qquad \text{(C15)}$$

Combining (C15) with (C1), we get the result (5.13) as was claimed. QED

---

* Supported in part by the National Science Foundation, Center for Theoretical Physics under Grant No. NSF GU 2061.

1 The singular character of nontempered strictly local fields is allowed according to general principles. See A. Jaffe, Phys. Rev. **158**, 1454 (1967).

2 K. Hepp, Les Houches Lectures (1970).

3 N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959).

4 Any of the rigorous renormalization schemes, e.g., Bogoliubov's (Ref. 3) or Speer's (Refs. 9, 10) go through for $:\phi^n:$ theory, any finite n.

5 H. Lehmann and K. Pohlmeyer, Commun. Math. Phys. **20**, 101 (1971).

6 See Ref. 3.

7 As formulated in Ref. 5, one demands that the space (or time) smeared part of the piece of the T product not fixed by unitarity and causality be continuous in the remaining variables which rules out the presence of distributions concentrated at the light cone vertex.

8 P. K. Mitter, Commun. Math. Phys. **20**, 251 (1971).

9 E. Speer, J. Math. Phys. **9**, 1404 (1968); *Generalized Feynman Amplitudes*, Annals of Math. Studies, No. 62 (Princeton U.P., Princeton, N.J., 1969).

10 E. Speer, "On the Structure of Analytic Renormalization," M.I.T. (Dept. of Mathematics), Preprint 1971.

11 Since the index of divergence $\omega(g) < 0$ for all the convergent graphs, Feynman amplitudes $\to 0$ for large momenta whence regularity on surfaces of coinciding prints can be shown to follow. This assumes that we do not make subtractions for convergent graphs.

12 E. Speer and M. J. Westwater, Ann. Inst. Henri Poincaré (A) **14**, 1 (1971).

13 K. Pohlmeyer, Institute for Advanced Study, preprint (1971). We received this preprint following the completion of the major part of this paper. Although the final expressions obtained by the two methods appear in different mathematical forms [c.f. the finite renormalization term $\tilde{\chi}(p)$ in our equation (5.2) with $\tilde{R}^-$ in Eq. (19) of Pohlmeyer] minimum singularity implies their equality.

14 Throughout this paper we use the theory of distributions depending on complex parameters. See Ref. 17, especially Appendix 2, pp. 147–51 and pp. 247–95.

15 *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. 1.

16 How to do this has been shown in Ref. 12, Lemmas 2.2.20, 2.2.33, and Theorem 2.3.1. It can be seen that a massless multiplet can be regulated with a single parameter, the massless GFA existing in a restricted region of regulating parameters. This motivates the simple regularization of multiplets, and we proceed to give consistence with the results of Ref. 12.

17 I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.

18 See Ref. 17, Chap. III, Secs. 2.3, 2.4 for quadratic forms raised to powers as distributions. Equation (6), p. 272 is useful for taking boundary values.

19 For the usual procedure for Eulerian integrals, see, e.g., Ref. 15, p. 60.

20 M. Daniel and P. K. Mitter (unpublished).

21 These have been formalized as axioms in Ref. 2, Chap. III. P.K.M. thanks R. F. Streater for a copy of Ref. 2.

22 We follow the method given in Ref. 2, Chap. IV. We emphasize that in Eqs. (4.17) and (4.18), the $\eta \to 0^+$ limit has already been taken.

23 It is possible, although laborious, to verify explicitly the cancellation mechanism at least for the first few terms.

24 It seems worthwhile to note that our procedure is analogous to what would have happened if we applied an analytic evaluation (Ref. 10), instead of a generalized evaluation (Ref. 9) as done in Ref. 2, to the generalized unitarity relation for a GFA. We also remark that an additive interpretation of our renormalization is available; for the three point massless GFA, the set of λ singularities as obtained from the sectorial representation is actually minimal, and from the work of Appendix B the subtraction terms are vertex parts.

25 The choice of $\tilde{\chi}(p)$, which may seem mysterious at this stage, is motivated by the work of Appendix C, where the real part of the double series in (5.2), in the region (5.9), is actually summed and the step from (5.12) and (5.13) is justified.

26 Reference 15, Vol. 2.

27 The choice of $\sin^2\pi z$ term was motivated by a suggestion of H. Lehmann.

28 In taking the real part we have $e^{i\pi z} \to \cos\pi z$, which combines with $\cot\pi z(1+\sin^2\pi z)$ to give $(1/\sin\pi z)(1-\sin^4\pi z)$ and terms involving $\sin^3\pi z$ give no contribution since they are analytic.

29 To get the asymptotic estimate, we divide the region of τ integration into two parts: $(0,(1/\zeta)^{2/3})$ and $((1/\zeta)^{2/3},\infty)$. In the second interval, the integral may be bounded, using Poisson's bound for Bessel functions, and we get the asserted decrease. In the first interval, for ζ sufficiently large, replace $J_1(2(1/\tau)^{1/2})$ by its asymptotic value; also the dominant contribution is for $t_1\zeta^{2/3}$ bounded, $t_1t_2\zeta^{2/3}$ bounded. After changing variables $\tau = \zeta^{-2/3}\xi$, a steepest descent estimate leads to the asserted decrease.

30 The necessity of obtaining a decrease property in region (ii) was pointed out to us by H. Lehmann (private communication). A similar argument is given in Ref. 13.

31 The method of Appendix C was developed in collaboration with D. I. Fivel.

32 See Ref. 15, p. 216, formula (4).

33 G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1958), 2nd ed., p. 79.

Thermodynamics and Hydrodynamics for a Modeled Fluid

J. Hardy and Y. Pomeau
Laboratoire de Physique des Plasmas, Faculté des Sciences, 91 Orsay, France*
(Received 6 August 1971)

This article presents a general study of a two-dimensional fluid model with microscopic discrete velocities. The rather unusual properties of this model lead to precise thermodynamical laws. The Navier-Stokes hydrodynamical equations are obtained, which contain a transport coefficient given by a Green-Kubo integral, and it is shown that this integral does not converge for a reason common to all two-dimensional fluids.

1. INTRODUCTION

By using the Green-Kubo method,[1] we are able to evaluate the transport coefficients of a fluid, such as diffusion coefficients, shear and bulk viscosities, thermal conductivity, etc., upon integrating the autocorrelation functions with respect to time.

It has been recently proved[2] that these autocorrelation functions behave like $t^{-\nu/2}$ for large values of t in the case of a ν-dimensional fluid; when $\nu = 2$, the usual hydrodynamical equations are no longer valid. For this reason, it seems legitimate to attempt further progress in hydrodynamics of two-dimensional fluids through computational experiments.

Unfortunately, due to limited capacity, the computer cannot simulate a fluid in a satisfactory manner for "realistic" models (with, for example, a Lennard-Jones or a hard-disc interaction between particles). In fact, the volume in phase space in which the system is located, grows with time, due to a structural instability of the many-body system. For this reason, the cumulative errors on the values of the autocorrelation functions do not allow us to know the asymptotic behavior of these functions.

This difficulty, inherent in any continuous model, may be partially removed in a model in which positions are distributed continuously, but where velocities are not; this is the case for the Maxwell model.

A precise definition of this model is, however, lacking and the connection between many-body dynamics and hydrodynamics is still obscure.

The purpose of this paper is to construct the thermodynamical and hydrodynamical theory for this fluid model. In Sec. 2, the Liouville equation is given by introducing a precise definition of the interaction laws. In Sec. 3, we shall describe the thermodynamical properties of the model and give an expression for the equilibrium distribution function.

The Green-Kubo method permits us to obtain successively the particular forms of the hydrodynamical equations: The Euler equations are given in Sec. 4 and the Navier-Stokes equations in Sec. 5.

Finally, Sec. 6 is devoted to the divergence problem of the Green-Kubo integral defining the transport coefficient of the model.

In conclusion, we shall examine the relevance of this model with more "physical" two-dimensional fluids.

2. DEFINITION OF THE MAXWELL MODEL

In molecular dynamics, a simple model is often used,[3] where particles interact with one another as hard spheres or hard discs. The exact dynamics are deduced from a simple algebra allowing us, in principle, to follow a phase space trajectory for a long time. But, since velocities are known only with limited precision, the domain of precision in phase space grows with time in these models, even in the free motion.

On the contrary, in a model with quantized velocities, the domain of precision remains constant in free motion and grows stepwise at each collision.

Maxwell[4] conceived such a model with quantized velocities. This model has been studied by several authors[5,6] from the viewpoint of the Boltzmann equation; but to our knowledge, a precise definition of the interaction laws is still lacking.

For computing the autocorrelation function from molecular dynamics, we must make precise the laws of dynamics for this Maxwell model:

(i) The Maxwell model is a two-dimensional "fluid" of N identical particles moving with any of the four unit velocities $(\vec{e}_{\mathrm{I}}, \vec{e}_{\mathrm{II}}, \vec{e}_{\mathrm{III}}, \vec{e}_{\mathrm{IV}})$ as shown in Fig. 1.

Owing to the discrete character of the velocity space, the individual trajectories trace broken lines, any break denoting a collision.

(ii) Collisions are binary and instantaneous.

(iii) Collisions may occur only between particles with opposite velocities (for example, $\vec{e}_{\mathrm{I}}$ and $\vec{e}_{\mathrm{III}}$).

(iv) In order to satisfy principles of mechanics (reversibility and conservation of momentum and energy), we are led to define the following interaction law.

A collision occurs between particles i and j $(i, j \in \{1, 2, \ldots, N\})$, when $(\vec{r}_i - \vec{r}_j)$ is parallel to the vector $\vec{e}_I + \vec{e}_{II}$ and $|\vec{r}_i - \vec{r}_j| < \sqrt{2}$ ($\vec{r}_i$ is the position vector of the ith particle).

The velocities of the two colliding particles change instantaneously.

Let $\vec{u}$ be the velocity of a particle before collision, then after collision it is given by

$$\overleftrightarrow{\Pi}\cdot\vec{u}, \quad \text{where } \overleftrightarrow{\pi} \text{ is the tensor,}$$
$$\overleftrightarrow{\Pi} \equiv \overleftrightarrow{1} - (\vec{e}_{\mathrm{II}} - \vec{e}_{\mathrm{I}})(\vec{e}_{\mathrm{II}} - \vec{e}_{\mathrm{I}}). \tag{2.1}$$

We illustrate this law in Fig. 2.

It can be readily verified that the interaction defined above satisfies the requirement of microreversibility.

Remarks: We have only defined head-on collisions:

(1) By definition, the cross section is equal to zero for particles moving perpendicularly to one another.

(2) When two particles move with the same velocity, they can run as near as desired without interacting.

(3) The particles move on a two-dimensional torus of periodicity L in the directions $(\vec{e}_{\mathrm{I}}, \vec{e}_{\mathrm{III}})$ and

$(\vec{e}_{\mathrm{II}}, \vec{e}_{\mathrm{IV}})$. Since the fluid does not interact with fixed boundaries it may have a nonzero total momentum.

Having given a precise definition of the interaction laws, we now write the Liouville equation, which assumes a rather unusual form in this Maxwell model. The N-body distribution function evolves discontinuously at phase-space points where a collision may occur.

This Liouville equation can be obtained by slightly modifying the Liouville equation for hard spheres.[7]

It reads

$$\frac{\partial D}{\partial t} + \sum_{K=1}^{N} \vec{u}_k \cdot \frac{\partial D}{\partial \vec{r}_k} = \sum_{i=1}^{N} \sum_{<j=2}^{N} \phi(\vec{r}_i - \vec{r}_j) T_{ij} D, \qquad (2.2)$$

where D is the N-body distribution function, $\vec{r}_k$ and $\vec{u}_k$ are the position and velocity vectors of the kth particle (recall $\vec{u}_k$ belongs to the finite set $\{\vec{e}_{\mathrm{I}}, \vec{e}_{\mathrm{II}}, \vec{e}_{\mathrm{III}}, \vec{e}_{\mathrm{IV}}\}$), where $\phi(\vec{r})$ is the Schwartz distribution defined for any continuous function $f(x, y)$ by the relation

$$\int \phi(\vec{r}) f(\vec{r}) d\vec{r} = \frac{1}{\sqrt{2}} \int_{-1/2}^{+1/2} f(x, x) dx, \qquad (2.3)$$

$\vec{r} = (x, y)$, T_{ij} is an operator on functions of $\vec{u}_i$ and $\vec{u}_j$:

$$T_{ij} g(\vec{u}_i, \vec{u}_j) \equiv \{g(\overleftrightarrow{\pi} \cdot u_i \overleftrightarrow{\pi} \cdot \vec{u}_j) - g(\vec{u}_i, \vec{u}_j)\} \delta_{\vec{u}_i, -\vec{u}_j}, \qquad (2.4)$$

where

$$S_{\vec{u}_i, -\vec{u}_j} = 0 \quad \text{if } \vec{u}_i + \vec{u}_j \neq 0,$$
$$\delta_{\vec{u}_i, -\vec{u}_j} = 1 \quad \text{if } \vec{u}_i + \vec{u}_j = 0.$$

With these mechanical properties of the model, we shall apply the Green–Kubo method to deduce its transport properties.

3. EQUILIBRIUM PROPERTIES

We shall study in this section the thermodynamics of the Maxwell model. As will be seen later, the usual methods of equilibrium thermodynamics are very convenient for this model, owing to the discrete character of the velocity space.

The main results of this Section are:

(1) at equilibrium, particles are uncorrelated;

(2) in the thermodynamical limit, we obtain

$$N_{\mathrm{I}} N_{\mathrm{III}} / N^2 = N_{\mathrm{II}} N_{\mathrm{IV}} / N^2;$$

N_α being the number of particles with a velocity equal to $\vec{e}_\alpha$ (α belongs to the finite set $\{\mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IV}\}$ and

$$N = \sum_{\alpha=\mathrm{I}}^{\alpha=\mathrm{IV}} N_\alpha \qquad (3.1)$$

stands for the total number of particles moving on the torus. We shall denote by $N\vec{w}^e$ the total momentum of the particles on the torus:

$$N_{\mathrm{I}} - N_{\mathrm{III}} = N w_x^e; \quad N_{\mathrm{II}} - N_{\mathrm{IV}} = N w_y^e. \qquad (3.2)$$

N, w_x^e, and w_y^e are constants of the motion.

We know an exact solution of the Liouville equation—the equilibrium Gibbs function which is the product of a constant density in configuration space $(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ and of a constant density H in velocity space. The latter, proportional to the number of different realizations for a given set $\{N_{\mathrm{I}}, N_{\mathrm{II}}, N_{\mathrm{III}}, N_{\mathrm{IV}}\}$ is given by

$$H(N_{\mathrm{I}}, N_{\mathrm{II}}, N_{\mathrm{III}}, N_{\mathrm{IV}}) = \frac{1}{Z(N, \vec{w}^e)} \times \frac{N!}{\Pi_{\alpha=\mathrm{I}}^{\mathrm{IV}} N_\alpha!}, \qquad (3.3)$$

where $Z(N, \vec{w}^e)$, a normalization constant, reads

$$Z(N, \vec{w}^e) = \sum_B H(N_{\mathrm{I}}, N_{\mathrm{II}}, N_{\mathrm{III}}, N_{\mathrm{IV}}).$$

The sum is taken over the set of the $\{N_{\mathrm{I}}, N_{\mathrm{II}}, N_{\mathrm{III}}, N_{\mathrm{IV}}\}$ satisfying (3.1) and (3.2).

Due to the existence of the constants of motion $(N, \vec{w}^e)$, H depends only on the discrete intensive parameter X:

$$X = (1/4N)(N_{\mathrm{I}} - N_{\mathrm{II}} + N_{\mathrm{III}} - N_{\mathrm{IV}}). \qquad (3.4)$$

The most probable state is realized when H is maximum, that is for a value $\langle X \rangle$ of X defined by

$$|H(\langle X \rangle + (1/N)) - H(\langle X \rangle)| = \min_X |H(X + (1/N)) - H(X)|. \qquad (3.5)$$

Note that X takes discrete values: $X = k/N$, where k is integer. From (3.3), we obtain

$$|N_{\mathrm{I}}(\langle X \rangle) N_{\mathrm{III}}(\langle X \rangle) - N_{\mathrm{II}}(\langle X \rangle) N_{\mathrm{IV}}(\langle X \rangle)| = \min_X |N_{\mathrm{I}}(X) N_{\mathrm{III}}(X) - N_{\mathrm{II}}(X) N_{\mathrm{IV}}(X)|. \qquad (3.6)$$

Taking account of the existence of the constants of

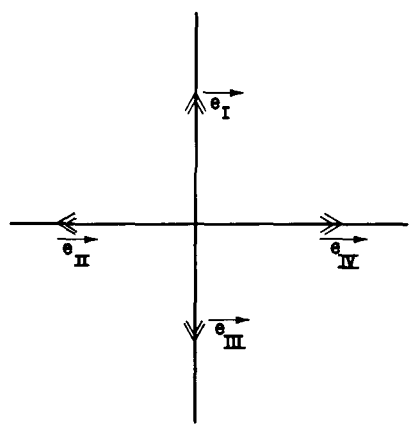

FIG. 1.

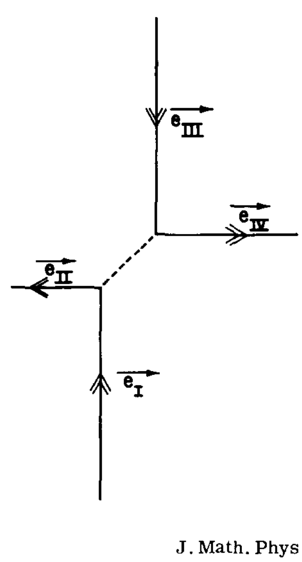

FIG. 2.

motion, $\langle X\rangle$ is given in the thermodynamical limit as follows:

$$\langle X\rangle \xrightarrow[N/L^2 \text{ fixed}]{N\to\infty} (w_x^e)^2 - (w_y^e)^2. \tag{3.7}$$

Using Stirling's formula, we show immediately from (3.3) that, at equilibrium, the mean square fluctuation of X satisfies

$$\langle\langle X - \langle X\rangle\rangle\rangle \sim 1/N. \tag{3.8}$$

As usual this fluctuation of the intensive parameter X is of order $n^{-1/2}$ in the thermodynamical limit. The quantity $\ln H$ is the so-called statistical entropy; it can not, however, be interpreted as the entropy defined from the Carnot principle, since, as shown in Appendix A, it does not satisfy, in the perfect fluid approximation, the equation

$$\left(\frac{\partial}{\partial t} + \vec{u}\cdot\vec{\nabla}\right)\ln H = 0$$

as is the case for the usual entropy.

At equilibrium, particles are uncorrelated and there exists a stationary and factorized solution D^e of the Liouville equation which may be identified as the canonical Gibbs equilibrium distribution function. It is

$$D^e(\{\vec{r}_i, \vec{u}_i\}) \equiv \frac{1}{N^N}\prod_{i=1}^{N} f^e(\vec{u}_i), \tag{3.9}$$

where $f^e(\vec{u}_i)$ is the one-body distribution function whose four values $\{f^e(\vec{e}_{\mathrm{I}}), f^e(\vec{e}_{\mathrm{II}}), f^e(\vec{e}_{\mathrm{III}}), f^e(\vec{e}_{\mathrm{IV}})\}$ are the components of a 4-vector denoted f^e. The function D^e is a stationary solution of the Liouville equation if we take

$$T_{ij}[f^e(\vec{u}_i)f^e(\vec{u}_j)] = 0 \tag{3.10}$$

for any value of i and j.

On the other hand, one has to choose the function f^e so that

$$\sum_{\alpha=\mathrm{I}}^{\mathrm{IV}} f^e(\vec{e}_\alpha) \equiv n^e, \tag{3.11a}$$

$$f^e(\vec{e}_{\mathrm{I}}) - f^e(\vec{e}_{\mathrm{III}}) \equiv n^e w_x^e, \tag{3.11b}$$

$$f^e(\vec{e}_{\mathrm{II}}) - f^e(\vec{e}_{\mathrm{IV}}) \equiv n^e x_y^e, \tag{3.11c}$$

where n^e is the density of particles

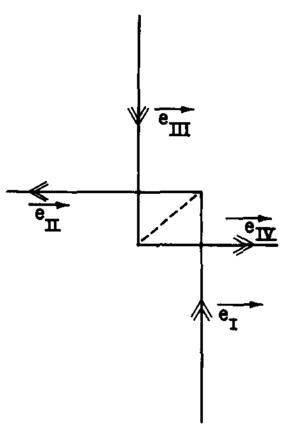

FIG. 3.

$$n^e \equiv N/L^2. \tag{3.12}$$

Equations (3.10) and (3.11) determine $f^e(\vec{u})$:

$$f^e(\vec{u}) = \tfrac{1}{4}n^e[1(\vec{w}^e)^2 + 2(\vec{w}^e\cdot\vec{u})^2 + 2\vec{w}^e\cdot\vec{u}], \tag{3.13}$$

where $\vec{u}$ is one of the four unit vectors $\{\vec{e}_{\mathrm{I}}, \vec{e}_{\mathrm{II}}, \vec{e}_{\mathrm{III}}, \vec{e}_{\mathrm{IV}}\}$.

For example, when $u = e_{\mathrm{I}}$ we have

$$f^e(\vec{e}_{\mathrm{I}}) = \tfrac{1}{4}n^e[1 + (w_x^e) - (w_y^e)^2 + 2w_x^e]. \tag{3.14}$$

Since $f^e(\vec{u})$ is positive definite, the mean velocity $\vec{w}^e$ lies inside the square defined by $\min_{\vec{u}}[f^e(\vec{u})] \geqslant 0$, where $\vec{u}$ is one of the four unit vectors $\{e_{\mathrm{I}}, e_{\mathrm{II}}, e_{\mathrm{III}}, e_{\mathrm{IV}}\}$. This square is drawn in Fig. 3.

Remark: The set $\{f^e(\vec{e}_{\mathrm{I}}), f^e(\vec{e}_{\mathrm{II}}), f^e(\vec{e}_{\mathrm{III}}), f^e(\vec{e}_{\mathrm{IV}})\}$ define a vector f^e in a four-dimensional space; these components, called "natural," are not simple and it is often advantageous to write f^e in another frame, called "hydrodynamical," where the three first components represent the hydrodynamical field, that is, the density and the mean velocity.

This "hydrodynamical" frame $\{s, p, q, z\}$ is defined from its corresponding components in the natural frame:

$$s = \tfrac{1}{4}(1, 1, 1, 1), \tag{3.15a}$$

$$p = \tfrac{1}{2}(1, 0, -1, 0), \tag{3.15b}$$

$$q = \tfrac{1}{2}(0, 1, 0, -1), \tag{3.15c}$$

$$z = \tfrac{1}{4}(1, -1, 1, -1). \tag{3.15d}$$

We now define a scalar product in the space of the one-body distribution functions

$$f \circ g \equiv \sum_{\alpha=\mathrm{I}}^{\mathrm{IV}} f(\vec{e}_\alpha)g(\vec{e}_\alpha), \tag{3.16}$$

where one finds, for example, the following expressions:

$$f \overset{e}{\circ} s = n^e, \tag{3.17a}$$

$$f \overset{e}{\circ} p = n^e w_x^e, \tag{3.17b}$$

$$f \overset{e}{\circ} q = n^e w_y^e, \tag{3.17c}$$

$$f \overset{e}{\circ} z = n^e[(w_x^e)^2 - (w_y^e)^2]. \tag{3.17d}$$

4. THE EULER HYDRODYNAMICAL EQUATIONS

In this section the hydrodynamical equations relating the number and momentum density are derived in the perfect fluid approximation (or Euler approximation).

Integrating (2.3) over $(N-1)$ positions and summing on the corresponding velocity variables, we get the first equation of the BBGKY hierarchy:

$$\frac{\partial f}{\partial t} + \vec{u}_1\cdot\frac{\partial f}{\partial \vec{r}_1} = \int d\vec{r}_2 d\vec{u}_2 \phi(\vec{r}_1 - \vec{r}_2)T_{12}g, \tag{4.1}$$

where

$$f(\vec{r}_1, \vec{u}_1) \equiv N\int d\vec{r}_2 d\vec{u}_2 \cdots d\vec{r}_N d\vec{u}_N D(\{\vec{r}_i, \vec{u}\}); \tag{4.2}$$

the integration over velocities is used here for simplicity, but as a matter of fact, $\int d\vec{u} f(\vec{u})$ means $\sum_{\alpha=\mathrm{I}}^{\mathrm{IV}} f(\vec{e}_\alpha)$ and

$$g(\vec{r}_1, \vec{u}_1 \cdot \vec{r}_2, \vec{u}_2) \equiv N(N-1) \times \int d\vec{r}_3 d\vec{u}_3 \cdots d\vec{r}_N d\vec{u}_N D(\{\vec{r}_i, \vec{u}_i\}). \quad (4.3)$$

The one-body distribution function permits us to express the number density and the local mean velocity through relations analogous to (3.17):

$$f(\vec{r},t) \circ s \equiv n(\vec{r},t), \quad (4.4a)$$

$$f(\vec{r},t) \circ p \equiv n(\vec{r},t) w_x(\vec{r},t), \quad (4.4b)$$

$$f(\vec{r},t) \circ q \equiv n(\vec{r},t) w_y(\vec{r},t). \quad (4.4c)$$

We can obtain the conservation equations by projecting Eq. (4.1) on the three hydrodynamical vectors (s,p,q) by means of the scalar product defined in (3.16) and the relations (4.4).

The conservation of the number density can be found with no further assumptions. In fact, the relation

$$S \circ \int d\vec{u}_2 T_{12}[g(\vec{u}_1, \vec{u}_2)] = 0 \quad (4.5)$$

together with the s component of both members of (4.1) yield

$$\frac{\partial (f \circ s)}{\partial t} + s \circ \left(\vec{u}_1 \cdot \frac{\partial f}{\partial \vec{r}_1}\right) = 0 \quad (4.6)$$

or, equivalently,

$$\frac{\partial n}{\partial t} + \vec{\nabla} \cdot (n\vec{\omega}) = 0. \quad (4.7)$$

The equation of conservation of the momentum density, however, cannot be explicitly given in terms of $\{n, \vec{w}\}$. In fact, to obtain this equation, the knowledge of the z component of the one-body distribution function and of the two-body distribution function is required. In the previous section, we have already found the stationary and homogeneous solution of the Liouville equation. This solution is, however, inadequate for the description of transport processes.

In the limit of weak gradients, we may approximate the exact density in phase space D by a local equilibrium distribution function D^0, whose form is similar to the canonical equilibrium distribution function D^e:

$$D^0 \equiv \frac{1}{N^N} \prod_{i=1}^{N} f^0[n(\vec{r}_i, t); \vec{w}(\vec{r}_i, t); \vec{u}_i], \quad (4.8)$$

where

$$f^0[n(\vec{r}_i, t); \vec{w}(\vec{r}_i, t); \vec{u}_i] \equiv \tfrac{1}{4} n(\vec{r}_i, t) \times 1 - w^2(\vec{r}_i, t) + 2[\vec{w}(\vec{r}_i, t) \cdot \vec{u}_i]^2 + 2\vec{w}(\vec{r}_i, t) \cdot \vec{u}_i \quad (4.9)$$

is entirely determined from the exact one-body distribution function through the relations

$$f^0(\vec{r}_i, t) \circ s \equiv f(\vec{r}_i, t) \circ s \equiv n(\vec{r}_i, t), \quad (4.10a)$$

$$f^0(\vec{r}_i, t) \circ p \equiv f(\vec{r}_i, t) \circ p \equiv n(\vec{r}_i, t) w_x(\vec{r}_i, t), \quad (4.10b)$$

$$f^0(\vec{r}_i, t) \circ q \equiv f(\vec{r}_i, t) \circ q \equiv n(\vec{r}_i, t) w_y(\vec{r}_i, t). \quad (4.10c)$$

The function $f^0(\vec{r}_i, t)$, called the one-body distribution function of local equilibrium, depends on time and position through the hydrodynamical field $\{n, w\}$.

The solution of the Liouville equation will be expanded around D^0 in powers of the gradients. In this expansion, we shall use an equivalence relation $a \stackrel{i}{=} b$ which means that the difference between a and b depends on time and space derivatives of $\{n, w\}$ of an order higher than i. It is obvious that, when i goes to infinity, this equivalence relation is reduced to a simple equality.

The one- and the two-body distribution functions read in the limit of the weak gradients:

$$f(\vec{r}_1; t; \vec{u}_1) \stackrel{0}{=} f^0(\vec{r}_1; t; \vec{u}_1), \quad (4.11a)$$

$$g(\vec{r}_1, \vec{u}_1; \vec{r}_2, \vec{u}_2; t) \stackrel{0}{=} f^0(\vec{r}_1; t; \vec{u}_1) f^0(\vec{r}_2; t; \vec{u}_2). \quad (4.11b)$$

Substituting (4.11) into (2.5), we obtain

$$T_{12} g(\vec{r}_1, \vec{u}_1; \vec{r}_2, \vec{u}_2) \stackrel{1}{=} \{f^0(\vec{r}_1, \overleftrightarrow{\Pi} \cdot \vec{u}_1) f^0(\vec{r}_2, \overleftrightarrow{\Pi} \cdot \vec{u}_2) - f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2)\} \delta_{\vec{u}_1, -\vec{u}_2} \quad (4.12)$$

We can then prove that the right-hand side of Eq. (4.1) is equal to zero up to the first order in the gradients. In fact, owing to the factor $\phi(\vec{r}_1 - \vec{r}_2)$, the integral on the right-hand side of Eq. (4.1) is carried out in a domain defined by $|\vec{r}_1 - \vec{r}_2| < \sqrt{2}$.

Since, in our model the unit is a microscopic length, we can expand $f^0(\vec{r}_2, \vec{u}_2)$ in Taylor series near $f^0(\vec{r}_1, \vec{u}_2)$, in the hydrodynamical limit where the scale of spatial variations of f^0 is much larger than unity.

Since to the zeroth order in the gradients (4.12) vanishes (D^0 being an exact solution of the Liouville equation to this order), the lowest term in the weak-gradient expansion of $T_{12}[g(\vec{r}_i, \vec{u}_i; \vec{r}_2, \vec{u}_2)]$ is odd in $(\vec{r}_1 - \vec{r}_2)$; thus to this order the right-hand side of Eq. (4.1) is equal to zero.

By projecting both members of Eq. (4.1) on the vectors p and q, and replacing f by its local equilibrium value, we obtain local equations of conservation of momentum, called Euler's hydrodynamical equations, where we retain terms up to first order in the gradients of n and $\vec{w}$:

$$\frac{\partial n\vec{w}}{\partial t} + \frac{1}{2} \vec{\nabla} \cdot \overleftrightarrow{P}_0 = 0, \quad (4.13a)$$

where

$$(P_0)_{xx} \stackrel{1}{=} n(1 + w_x^2 - w_y^2), \quad (4.13b)$$

$$(P_0)_{yy} \stackrel{1}{=} n(1 + w_y^2 - w_x^2), \quad (4.13c)$$

$$(P_0)_{xy} \stackrel{1}{=} (P_0)_{yx} \stackrel{1}{=} 0. \quad (4.13d)$$

The Euler equations of hydrodynamics do not include any term similar to an intermolecular pressure. They introduce no irreversible transport phenomena. These transport phenomena will be accounted for at the next order in the Chapman–Enskog expansion: This will be done in the following section.

5. THE NAVIER–STOKES HYDRODYNAMICAL EQUATIONS

Using the Green–Kubo method, we shall find in this section the so-called Navier–Stokes equations of hydrodynamics, which take irreversible processes into account. These equations include a transport coefficient given by the Green–Kubo integral. Before proceeding on to the calculations, let us explain the general features of the method.

The distribution function D is written as the sum of the local equilibrium distribution function D^0 and a function D^1; the Liouville equation allows us to calculate dD^1/dt, and upon integrating this equation by the method of characteristics we obtain $D \equiv D^0 + D^1$; substituting this expression into Eqs. (4. 2) and (4. 1), it will be possible to write the hydrodynamical components of Eq. (4. 1) which are just the conservation equations. In this procedure the operations of differentiation and integration can not be performed exactly. They can be done, however, with sufficient accuracy to give conservation equations correct to second order in the gradients of the hydrodynamical field $\{n, \vec{w}\}$.

To first order in the gradients of the hydrodynamical field, the Liouville equation reads

$$\frac{dD}{dt} \equiv \left(\frac{dD^0}{dt}\right)_k + \left(\frac{dD^0}{dt}\right)_U + \frac{dD^1}{dt} \stackrel{1}{=} 0, \tag{5.1}$$

where

$$\left(\frac{dD^0}{dt}\right)_k = D^0 \sum_{i=1}^{N} \frac{(\partial/\partial t) f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i] + \vec{u}_i \cdot (\partial/\partial \vec{r}_i) f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i]}{f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i]} \tag{5.2}$$

and where

$$\left(\frac{dD^0}{dt}\right)_U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\phi(\vec{r}_i - \vec{r}_j) T_{ij} \{f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i] f^0[n(\vec{r}_j, t) \vec{w}(\vec{r}_j, t) \vec{u}_j]\}}{f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i] f^0[n(\vec{r}_j, t), \vec{w}(\vec{r}_j, t); \vec{u}_j]}. \tag{5.3}$$

As is shown in Appendix C, the time derivative $(\partial/\partial t) f^0(n(r_i, t), w(r_i, t); u_i)$ appearing in (5. 2) can be calculated in the weak-gradient limit, by means of the Euler hydrodynamical equations

$$\left[\frac{dD^0}{dt}\right]_k \stackrel{1}{=} \frac{D^0}{2} \sum_{i=1}^{N} \frac{n^e \chi(\vec{u}_i)}{f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i]} \times \left[(1 - \vec{w}^2)\left(\frac{\partial w_x}{\partial x} - \frac{\partial w_y}{\partial y}\right) + 2 w_x w_y \left(\frac{\partial w_y}{\partial x} - \frac{\partial w_x}{\partial y}\right)\right], \tag{5.4a}$$

where

$$\chi(\vec{e}_\alpha) \equiv 1 - 2\vec{e}_\alpha \cdot \vec{e}_{\mathrm{II}} \tag{5.4b}$$

being one of the four indices I, II, III, IV. The "collision term" $[dD^0/dt]_U$ defined in (5. 3) is equal to zero to first order in the gradients of the hydrodynamical field n, w. In fact, we have

$$T_{ij}\{f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i] f^0[n(\vec{r}_j, t), \vec{w}(\vec{r}_j, t); \vec{u}_j]\} \stackrel{1}{=} T_{ij}\left(f^0[n(\vec{r}_i, t), \vec{w}(\vec{r}_i, t); \vec{u}_i](\vec{r}_i - \vec{r}_j) \cdot \frac{\partial}{\partial \vec{r}_i} f^0[n(\vec{r}\ , t), \vec{w}(\vec{r}_i, t); \vec{u}_j]\right). \tag{5.5}$$

Since the $i - j$ term in the right-hand side of the above equation is antisymmetric for interchange of i and j, we obtain

$$\left[\frac{dD^0}{dt}\right]_U \stackrel{1}{=} 0. \tag{5.6}$$

From (5. 6) and (5. 1), we obtain

$$\left[\frac{dD^0}{dt}\right]_k \stackrel{1}{=} -\frac{dD^1}{dt} \tag{5.7}$$

and then upon integrating along the trajectories

$$D^1 \stackrel{1}{=} -\int_{-\infty}^{t} \left[\frac{dD^0}{dt}\right]_k (t') dt'. \tag{5.8}$$

The integration in (5. 8) is carried out on points of the trajectory in phase space which lead to the point argument of D^1 at time t. Due to the time derivative of the integrand in the right-hand side of (5. 8), the integration can be performed up to first order in the hydrodynamical gradients by neglecting the dependence of $\{n, \vec{w}\}$ with respect to the microscopic time scale in which the integrand of (5. 8) goes to zero almost everywhere:

$$D^1 \stackrel{1}{=} \frac{D^0}{2}\left[(1 - \vec{w}^2)\left(\frac{\partial w_x}{\partial x} - \frac{\partial w_y}{\partial y}\right) + 2 w_x w_y \left(\frac{\partial w_y}{\partial x} - \frac{\partial w_x}{\partial y}\right)\right] \times \sum_{j=1}^{N} \int_{-\infty}^{t} \frac{\chi[\vec{u}_j(t')] n[\vec{r}_j, t']}{f^0[\vec{u}_j(t')]} dt'. \tag{5.9}$$

By inserting

$$D \stackrel{1}{=} D^0 + D^1 \tag{5.10}$$

into the definitions (4. 2) and (4. 3), we obtain

$$f(\vec{r}_1, \vec{u}_1) \stackrel{1}{=} f^0(\vec{r}_1, \vec{u}_1) + \frac{n^2}{2}\left[(1 - \vec{w}^2)\left(\frac{\partial w_x}{\partial x} - \frac{\partial w_y}{\partial y}\right) + 2 w_x w_y \left(\frac{\partial w_y}{\partial x} - \frac{\partial w_x}{\partial y}\right)\right] \times \int d\vec{r}_1 d\vec{r}_2 d\vec{u}_2 \cdots d\vec{r}_N d\vec{u}_N D^0(\{\vec{r}_i, \vec{u}_i\}) \times \sum_{j=1}^{N} \int_{-\infty}^{t} \frac{\chi[\vec{u}_j(t')] dt'}{f^0[\vec{r}_j(t'), \vec{u}_j(t')]} \tag{5.11}$$

and

$$g(\vec{r}_1, \vec{u}_1; \vec{r}_2, \vec{u}_2) \stackrel{1}{=} f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2) + \frac{n^3}{2}\left[(1 - \vec{w}^2)\left(\frac{\partial w_x}{\partial x} - \frac{\partial w_y}{\partial y}\right) + 2 w_x w_y \left(\frac{\partial w_y}{\partial x} - \frac{\partial w_x}{\partial y}\right)\right] \times \int d\vec{r}_1 d\vec{r}_2 d\vec{r}_3 d\vec{u}_3 \cdots d\vec{r}_N d\vec{u}_N D^0 \times (\{\vec{r}_i, \vec{u}_i\}) \sum_{j=1}^{N} \int_{-\infty}^{t} \frac{\chi[\vec{u}_j(t')] dt'}{f^0[\vec{r}_i(t'), \vec{u}_j(t')]}. \tag{5.12}$$

The first equation of the BBGKY hierarchy is then obtained up to second order in the gradients of the hydrodynamical field n, w; the p and q components of this equation give the Navier–Stokes equations of hydrodynamics:

$$\frac{\partial n\vec{w}}{\partial t} + \tfrac{1}{2}\vec{\nabla}\cdot\overleftrightarrow{P} + \vec{C} \stackrel{2}{=} 0, \tag{5.13}$$

where P stands for the pressure tensor $\overleftrightarrow{P} = \overleftrightarrow{P}_0 + \overleftrightarrow{P}_1$, $\overleftrightarrow{P}_0$ being the equilibrium pressure tensor defined in (4.12) and $\overleftrightarrow{P}_1$ the "viscous" pressure tensor given by

$$\overleftrightarrow{P}_1 \equiv \theta\left[(1-\vec{w}^2)\left(\frac{\partial w_x}{\partial x} - \frac{\partial w_y}{\partial y}\right) + 2w_x w_y\left(\frac{\partial w_y}{\partial x} - \frac{\partial w_x}{\partial y}\right)\right]\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{5.14}$$

The transport coefficient θ, analogous to a viscosity, is defined by the Green–Kubo integral

$$\theta \equiv \frac{(n^e)^2}{4}\int_0^\infty dt\left\langle \chi[\vec{u}_1(t=0)]\sum_{i=1}^{N}\frac{\chi[\vec{u}_i(t)]dt'}{f^0[\vec{r}_i(t'),\vec{u}_i(t')]}\right\rangle, \tag{5.15}$$

where $\langle\ \rangle_0$ means as usual an equilibrium average over the ensemble of initial conditions.

The momentum flux C in Eq. (5.13) comes from the right-hand side of Eq. (4.1); when we calculate up to second order in the hydrodynamical gradients. The components of C (see Appendix D) read

$$C_x = \frac{n^2}{24}\left[w_x\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2(w_x^2 - w_y^2) - (1 + w_x^2 - w_y^2)\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 w_x\right] \tag{5.16a}$$

and

$$C_y = \frac{n^2}{24}\left[w_y\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2(w_y^2 - w_x^2) - (1 + w_y^2 - w_x^2)\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 w_y\right]. \tag{5.16b}$$

As anticipated, the transport coefficient θ, as defined by (5.15), is an integral over time of the autocorrelation function of

$$\chi(\vec{u})/f^0(\vec{u}). \tag{5.17}$$

The practical advantage of (5.15) is to permit the calculation of the transport coefficient θ from the autocorrelation function for a homogeneous fluid. The value of the transport coefficient depends not only on n but also on $\vec{w}$, due to the absence of Galilean invariance of the Maxwell model. In the small velocity limit $\vec{w}^2 \ll 1$, (5.15) becomes

$$\theta = n^e\int_0^\infty dt\left\langle \chi[\vec{u}_1(t=0)]\sum_{i=1}^{N}\chi[\vec{u}_i(t)]\right\rangle_0. \tag{5.18}$$

The Navier–Stokes equations thus allow us to study several properties of the Maxwell fluid. This is the purpose of the following section.

6. NORMAL MODES OF THE NAVIER–STOKES EQUATIONS

We shall study in this section the solutions of the Navier–Stokes equations linearized around the equilibrium state $\{n^e; w_x^e = 0, w_y^e = 0\}$.

Given a space-dependent initial condition of the hydrodynamical field, three modes evolve independently: one mode of vorticity diffusion which describes the evolution of divergence-free velocity fields and two sound waves describing the coupled evolution of the density and of an irrotationnal velocity field. The perturbations of the equilibrium hydrodynamical field are incorporated in the vector X:

$$X(\vec{r},t) \equiv (\delta n(\vec{r},t); w_x(\vec{r},t); w_y(\vec{r},t)) \tag{6.1}$$

with

$$\delta n \equiv (n - n^e)/n^e.$$

It is convenient to use the Fourier transform of the hydrodynamical field

$$\tilde{X}(\vec{k},t) \equiv \int d\vec{\rho}\, e^{2i\vec{k}\cdot\vec{\rho}}X(\vec{\rho},t) \tag{6.2}$$

to write the linearized Navier–Stokes equations in the form of a linear differential equation

$$M(\vec{k})\tilde{\chi}(\vec{k},t) + \frac{\partial\tilde{\chi}(\vec{k},t)}{\partial t} = 0, \tag{6.3}$$

where $M(\vec{k})$ is the 3×3 matrix:

$$M = \begin{bmatrix} 0 & 2ik\cos\zeta & 2ik\sin\zeta \\ ik\cos\zeta & 2k^2[(\theta/n)\cos^2\zeta + \frac{1}{12}n(\cos\zeta+\sin\zeta)^2] & -2k^2(\theta/n)\cos\zeta\sin\zeta \\ ik\sin\zeta & -2k^2(\theta/n)\cos\zeta\sin\zeta & 2k^2[(\theta/n)\sin^2\zeta + \frac{1}{12}n(\sin\zeta+\cos\zeta)^2] \end{bmatrix}$$

and $\vec{k} = k(\cos\zeta, \sin\zeta)$.

Equation (6.3) may be solved by expanding the vector

$$\tilde{X}(\vec{k},t=0) = \sum_{\mu=1}^{3} X^\mu(\vec{k})Y_\mu(\vec{k};t=0), \tag{6.4}$$

from which it follows immediately that

$$\tilde{X}(\vec{k},t) = \sum_{\mu=1}^{3} X^\mu(\vec{k})Y_\mu(\vec{k},t)$$

$$= \sum_{\mu=1}^{3} X^\mu(\vec{k})Y_\mu(\vec{k},t=0)e^{-t\omega_\mu(\vec{k})}, \tag{6.5}$$

where ω_μ is the eigenvalue of $Y_\mu(\vec{k},t)$. As usual these hydrodynamical eigenvalues satisfy $\omega_\mu(\vec{k}=\vec{0}) = 0$.

To each of the sound waves there correspond two eigenvectors

$$Y_{1,2}(\vec{k},t=0) = ((-1)^{1,2}\sqrt{2}; \cos\zeta; \sin\zeta) \tag{6.6a}$$

with the corresponding eigenvalues

$$\omega_{1,2}(\vec{k}) = i(-1)^{1,2}\sqrt{2} + k^2[(\theta/n)\cos^2 2\zeta + \tfrac{1}{12}n(1+\sin 2\zeta)] + O(k^3). \quad (6.6b)$$

The third mode of the perturbed hydrodynamical field is the mode of vorticity diffusion,

$$Y_3(\vec{k}, t=0) = (0, -\sin\zeta, \cos\zeta) \quad (6.6c)$$

with the eigenvalue

$$\omega_3(\vec{k}) = 2k^2[(\theta/n)\sin^2 2\zeta + \tfrac{1}{12}n(1+\sin 2\zeta)] + O(k^3) \quad (6.6d)$$

The results (6.6) validate *a posteriori* the linearization of the Navier–Stokes equations, since perturbations of small amplitude of the hydrodynamical field are damped. Although the Navier–Stokes equations have been obtained along the usual lines, we shall see in the next section that the Green–Kubo integral (5.18) does not converge for large time.

7. ASYMPTOTIC BEHAVIOR OF THE AUTOCORRELATION FUNCTION

In this section we apply a more general method to the Maxwell model.

It has been shown that the Green–Kubo integrands behave like $t^{-\nu/2}$ (ν = dimensionality) for large time, so that their integrals diverge for $\nu = 2$. This divergence does not rule out the Maxwell model.

Calculations are, however, a little different because of its anisotropy and of the absence of Galilean invariance.

We have seen in the fifth section that the transport coefficient θ depends on n and $\vec{w}$; to see this, we shall calculate the asymptotic behavior of the Green–Kubo integrand $\psi(t)$ for $\vec{w}^2 \ll 1$.

The transport coefficient θ is expressed as the time integral of

$$\psi(t) \equiv \left\langle \chi[\vec{u}_1(t=0)] \sum_{i=1}^{N} \chi[\vec{u}_j(t)] \right\rangle_0. \quad (7.1)$$

The problem of calculation of the autocorrelation function is tantamount to the solution of the Liouville equation with a given initial condition. The time dependence of $\sum_{i=1}^{N}\chi[\vec{u}_j(t)]$ can be expressed in terms of a solution $\hat{D}(\{\vec{\rho}_i, \vec{v}_i\}; t; \vec{r}_1, \vec{u}_1)$ of the Liouville equation with respect to the variables $\{\vec{\rho}_i; \vec{v}_i\}$:

$$\sum_{i=1}^{N} \chi[\vec{u}_i(t)] = \int d\vec{\rho}_1 d\vec{v}_1 \cdots d\vec{\rho}_N d\vec{v}_N \times \hat{D}(\{\vec{\rho}_i, \vec{v}_i\}; t; \vec{r}_1, \vec{u}_1) \sum_{j=1}^{N} \chi[\vec{v}_j]. \quad (7.2)$$

$\hat{D}$ satisfies the initial condition

$$\hat{D}(\{\vec{\rho}_i, \vec{v}_i\}; t=0; \vec{r}_1, \vec{u}_1) \equiv \frac{1}{N} D^e(\{\vec{v}_j\}) \sum_{j=1}^{N} \delta(\vec{r}_1 - \vec{\rho}_j)\delta(\vec{u}_1 - v_j), \quad (7.3)$$

in which D^e denotes the canonical equilibrium distribution function. Since $\hat{D}$ is an exact solution of the Liouville equation, symmetrical with respect to the interchange of particles, $\hat{D}$ has the usual properties of the N-body distribution function. In particular we can write

$$\hat{D}(\{\vec{\rho}_i, \vec{v}_i\}; t; \vec{r}_1, \vec{u}_1) \equiv \hat{D}^0(\{\vec{\rho}_i, \vec{v}_i\}; t; \vec{r}_1, \vec{u}_1) + \hat{D}^1(\{\vec{\rho}_i, \vec{v}_i; t; \vec{r}_1, \vec{u}_1), \quad (7.4)$$

where $\hat{D}^0$ is a local equilibrium distribution function defined by

$$\hat{D}^0(\{\vec{\rho}_i, \vec{v}_i\}; t; \vec{r}_1, \vec{u}_1) \equiv \frac{1}{N^N} \prod_{i=1}^{N} \frac{n(\vec{\rho}_i; \vec{r}_i, \vec{u}_1)}{4} \{1 - \vec{w}^2(\vec{\rho}_i; \vec{r}_1, \vec{u}_1) + 2[\vec{w}(\vec{\rho}_i; \vec{r}_1, \vec{u}_1)\cdot \vec{v}_i]^2 + 2\vec{w}(\vec{\rho}_i; \vec{r}_1, \vec{u}_1)\cdot \vec{v}_i\}, \quad (7.5)$$

where

$$n(\vec{\rho}_i; \vec{r}_1, \vec{u}_1) \equiv \int d\vec{\rho}_1' d\vec{v}_1' \cdots d\vec{\rho}_N' d\vec{v}_N' \hat{D} \times (\{\vec{\rho}_i', \vec{v}_i'\}; t; \vec{r}_1, \vec{u}_1)\delta(\vec{\rho}_i' - \vec{\rho}_1') \quad (7.6)$$

and

$$n(\vec{\rho}_i; \vec{r}_1, \vec{u}_1)\vec{w}(\vec{\rho}_i; \vec{r}_1, \vec{u}_1) \equiv \int d\vec{\rho}_1' d\vec{v}_1' \cdots d\vec{\rho}_N' d\vec{v}_N' \hat{D}(\{\vec{\rho}_i', \vec{v}_i'\}; t; \vec{r}_1, \vec{u}_1) \times \delta(\vec{\rho}_i' - \vec{\rho}_1')\vec{v}_1'. \quad (7.7)$$

Substituting (7.4) into Eq. (7.2), it yields

$$\psi(t) = \psi^0(t) + \psi^1(t) \quad (7.8)$$

with

$$\psi^0(t) = \langle \chi[\vec{u}_1(t=0)] \int d\vec{\rho} n(\vec{\rho}; t; \vec{r}_1, \vec{u}_1) \times (w_x^2 - w_y^2)(\vec{\rho}; t; \vec{r}_1, \vec{u}_1)\rangle_0. \quad (7.9)$$

An asymptotic value of $\psi^0(t)$ can now be calculated from the value of the hydrodynamical field $\{n, \vec{w}\}$ at $t \to \infty$. In fact we know that, for large time, the intensive hydrodynamical fields describing, respectively, the local equilibrium distribution function $\hat{D}^0$ and the canonical equilibrium distribution function D^e are found to be the same.

The calculation of the long time behavior of the hydrodynamical field can then be carried out by using the laws of linearized hydrodynamics about the equilibrium hydrodynamical field $\{n^e; w_x^e = 0; w_y^e = 0\}$. By expanding the Fourier-transform of the mean velocity on the basis of the eigenvectors $Y_\mu(\vec{k}, t)$ of the evolution matrix M, we obtain

$$(\tilde{w}_x^2 - \tilde{w}_y^2)(\vec{k}, t) \xrightarrow{t\to\infty} (\cos^2\zeta - \sin^2\zeta) \times [\hat{w}_x(\vec{k}, t=0)\cos\zeta + \hat{w}_y(\vec{k}, t=0)\sin\zeta]^2 \times \exp\{-2k^2 t[(\theta/n)\cos^2 2\zeta + \tfrac{1}{12}n(1+\sin 2\zeta)]\} + [-\tilde{w}_x(\vec{k}, t=0)\sin\zeta + \tilde{w}_y(\vec{k}, t=0)\cos\zeta]^2 \times \exp\{-4k^2 t[(\theta/n)\sin^2 2\zeta + \tfrac{1}{12}n(1+\sin 2\zeta)]\}. \quad (7.10)$$

The integral of the right-hand side of Eq. (7.9) can be evaluated from the Plancherel–Parseval theorem

$$\int d\vec{\rho} n(\vec{\rho}; t; \vec{r}_1, \vec{u}_1)(w_x^2 - w_y^2)(\vec{\rho}; t; r_1, u_1) \xrightarrow{t\to\infty} n^e(\hat{w}_x^2 - w_y^2)(k=0; t; \vec{r}_1, \vec{u}_1)(A/t), \quad (7.11)$$

where A is the sum of the two integrals

$$\frac{1}{4\pi^2}\int_0^{2\Pi} \frac{\cos^2 2\zeta\, d\zeta}{(\theta/n)\cos^2 2\zeta + \frac{1}{12}n(1+\sin^2\zeta)} = \frac{n}{2\pi\theta}\left[1 - \left(\frac{n^2}{n^2 + 24\theta}\right)^{1/2}\right], \quad (7.12a)$$

and

$$\frac{1}{8\pi^2}\int_0^{2\pi}\frac{\cos^2 2\zeta d\zeta}{(\theta/n)\sin^2 2\zeta+\frac{1}{12}n(1+\sin 2\zeta)}=\frac{n}{4\pi\theta}\left[-1+\left(\frac{n^4-24\theta n^2-288\theta^2-24\theta(144\theta^2+24n^2\theta)^{1/2}}{n^2(n^2-48\theta)}\right)^{1/2}\right]. \tag{7.12b}$$

By means of the conservation of the total momentum defined by

$$\begin{aligned}\int d\vec{\rho}\, n(\vec{\rho};t;\vec{r}_1,\vec{u}_1)\vec{w}(\vec{\rho};t;\vec{r}_1,\vec{u}_1) &= \int d\vec{\rho}_1 d\vec{v}_1\cdots d\vec{\rho}_N d\vec{v}_N \vec{v}_1\\ &\times \hat{D}(\{\vec{\rho}_i,\vec{v}_i\};t=0;\vec{r}_1,\vec{u}_1)\end{aligned} \tag{7.13}$$

and of the explicit expression for $\hat{D}(\{\vec{\rho}_i,\vec{v}_i\};t=0;\vec{r}_1,\vec{u}_1)$ given by (7.3), we obtain

$$\begin{aligned}\int d\vec{\rho}\,\vec{w}(\vec{\rho};t;\vec{r}_1,\vec{u}_1) &\equiv \tilde{w}(\vec{k}\\ &=0;t;\vec{r}_1,\vec{u}_1)\xrightarrow{t\to\infty}(\vec{u}_1(t=0))/n^e.\end{aligned} \tag{7.14}$$

By inserting (7.14) into (7.11) we finally have

$$\psi^0(t)\xrightarrow{t\to\infty} A/n^e t. \tag{7.15}$$

The transport coefficient θ is given by the time integral of $n\{\psi^0(t)+\psi^1(t)\}$. We assume that $\psi^1(t)$ decay faster than $\psi^0(t)$ for large time (this assumption is the basic point of the Bogulyubov theory).

This result shows that the calculations of Sec. 5 are not consistent, even if we assume that $\theta=0$. In the conservation equations, the instantaneous variation of momentum cannot be related to the divergence of a sort of local viscous pressure.

Once the Maxwell model is defined with a view to numerical simulation, it should be easy to check directly the asymptotic behavior of $\psi(t)$ for large time.

8. CONCLUSION

Because of its anisotropy, the Maxwell model exhibits two particularities:

(i) the transport coefficient θ depends not only on density but on the mean velocity $\vec{w}$.

(ii) In the Navier–Stokes equations, the momentum flux cannot be written in the form of a pressure divergence.

The advantage of the Maxwell model lies in its dynamics being well-adapted to numerical computations of the long-time behavior of the autocorrelation function $\psi^0(r)$. Although this model has a very particular property, it presents the same kind of divergence as in more "physical" two-dimensional fluids. Here by "physical" we mean that the particles interact through a continuous potential.

APPENDIX A: NONCONSERVATION OF ENTROPY IN PERFECT FLUID FLOWS

The adiabaticity of a flow is characterized by the conservation of entropy during the motion. More precisely this property is described by the equation

$$DS=0, \tag{A1}$$

where S is the entropy per particle and D is the material derivative operator $D\equiv[(\partial/\partial t)+\vec{u}\cdot\vec{\nabla}]$.

In classical fluids, it can be demonstrated that the entropy per particle is conserved in the approximation of the Euler equations. In the case of the Maxwell fluid, the anisotropy does not allow such a proof. In fact, assume that the entropy exists. Then the entropy depends on thermodynamical variables

$$S=S(n,w_x,w_y). \tag{A2}$$

Let us write

$$DS\equiv\frac{\partial S}{\partial n}Dn+\frac{\partial S}{\partial w_x}Dw_x+\frac{\partial S}{\partial w_y}Dw_y. \tag{A3}$$

Then the Euler equations enable us to express Dn, Dw_x, and Dw_y in terms of $(\partial w_x/\partial x)$, $(\partial w_y/\partial y)$, $(\partial w_y/\partial x)$, $(\partial w_y/\partial y)$. We find furthermore that Dn, Dw_x, and Dw_y are different from zero. The coefficients of $\partial w_x/\partial x$, $\partial w_x/\partial y$, $\partial w_y/\partial x$, and $\partial w_y/\partial y$ have to vanish in order to have $DS=0$. It is in general impossible (except for $w_x=w_y$) to have four independent equations with three unknowns. DS is thus different from zero. This proof holds when $S\equiv\ln H$, where H is defined in (3.3). The entropy is then not conserved in the Maxwell model in the approximation of the Euler equations.

APPENDIX B: THE ALGEBRA OF THE MAXWELL MODEL

In the course of calculations, we have to express hydrodynamical components of quantities which are products of functions of $\vec{u}$; for example, $\vec{u}\cdot(\partial f(\vec{u})/\partial\vec{r})$, where $\vec{u}$ is one of the four unit vectors $(\vec{e}_{\rm I},\vec{e}_{\rm II},\vec{e}_{\rm III},\vec{e}_{\rm IV})$.

Instead of evaluating hydrodynamical components of the product in each particular case, it is more convenient to give a general expression. We define for this purpose an internal product

$$c=a|b=b|a, \tag{B1}$$

where $c_\alpha=a_\alpha b_\alpha$ for any value of $\alpha=(\mathrm{I},\mathrm{II},\mathrm{III},\mathrm{IV})$.

Thus $\vec{u}\cdot(\partial f(\vec{u})/\partial\vec{r})$ reads $\vec{u}\cdot(\partial/\partial\vec{r})|f$. We have found the hydrodynamical components of such a product:

$$\begin{aligned}4c\circ s&=(a\circ s)(b\circ s)+(a\circ z)(b\circ z)\\&\quad+2[(a\circ p)(b\circ p)+(a\circ q)(b\circ q)],\end{aligned} \tag{B2a}$$

$$4c\circ p=(a\circ s+a\circ z)(b\circ p)+(a\circ p)(b\circ s+b\circ z), \tag{B2b}$$

$$4c\circ q=(a\circ s-a\circ z)(b\circ q)+(a\circ q)(b\circ s-b\circ z), \tag{B2c}$$

$$4c \circ z = (a \circ s)(b \circ z) + (a \circ z)(b \circ s) + 2[(a \circ p)(b \circ p) - (a \circ q)(b \circ q)]. \quad \text{(B2d)}$$

APPENDIX C: ELIMINATION OF EXPLICIT TIME DERIVATIVE IN $\left|\frac{dD^0}{dt}\right|_k$

By definition (6.5) we have

$$\left|\frac{dD^0}{dt}\right|_k = D^0 \sum_{i=1}^{N} \frac{1}{f^0(\vec{u}_i)}\left(\frac{\partial f^0(\vec{u}_i)}{\partial t} + \vec{u}_i \cdot \frac{\partial f^0(\vec{u}_i)}{\partial \vec{r}}\right). \quad \text{(C1)}$$

The Euler hydrodynamical equations are obtained by eliminating the purely hydrodynamical components of df^0/dt, i.e., $(df^0/dt) \circ s \overset{1}{=} 0$, $(df^0/dt) \circ p \overset{1}{=} 0$, and $(df^0/dt) \circ q \overset{1}{=} 0$. Therefore df^0/dt is "parallel" to the z vector; this procedure enables us to obtain the following identity:

$$z \circ \left(\frac{df^0}{dt}\right) \overset{1}{=} \left(4z \left| \frac{f^0(-w)}{f^0 \circ s}\right.\right) \circ \left(\frac{df^0}{dt}\right) \quad \text{(C2)}$$

since

$$z \circ \left(4z \left| \frac{f^0(-w)}{f^0 \circ s}\right.\right) \equiv 1. \quad \text{(C3)}$$

The internal product defined in Appendix B satisfies the identity

$$(a|b) \circ c = a \circ (b|c) = (a|c) \circ b = \sum_{\alpha=C}^{N} a_\alpha b_\alpha c_\alpha. \quad \text{(C4)}$$

This property allows us to express (C2) in the form

$$z \circ \left(\frac{df^0}{dt}\right) \overset{1}{=} 4\left(\frac{f^0(-w)}{-f^0 \circ s} \left| \frac{df^0(w)}{dt}\right.\right) \circ z. \quad \text{(C5)}$$

The equilibrium condition (3.6) is equivalent, in our formalism, to the statement that $f^0(w) | f^0(-w)$ is "parallel" to the s vector; the explicit time derivatives are then eliminated in (C5), since $s \circ z = 0$, and we find

$$\left[\frac{dD^0}{dt}\right]_k \overset{1}{=} D^0 \sum_{i=1}^{N} \frac{4\chi(\vec{u}_i)}{f^0(\vec{u}_i)} z \circ \left(\frac{f^0(-w)}{f^0 \circ s} \left| \frac{df^0(w)}{dt}\right.\right), \quad \text{(C6)}$$

where $\chi(\vec{u}_i)$ has already been used in (5.4b).

Performing scalar and internal products in (C6), we obtain

$$\left[\frac{dD^0}{dt}\right]_k = D^0 \sum_{i=1}^{N} \frac{\chi[\vec{u}_i]}{f^0(\vec{u}_i)}\left[\frac{f^0 \circ s}{2f^0 \circ s}\left(\frac{\partial f^0 \circ p}{\partial x} - \frac{\partial f^0 \circ q}{\partial y}\right) + \frac{2f^0 \circ z}{2f^0 \circ s}\left(\frac{\partial f^0 \circ p}{\partial x} + \frac{\partial f^0 \circ q}{\partial y}\right) - \frac{f^0 \circ p}{2f^0 \circ s}\frac{\partial}{\partial x}(f^0 \circ s + f^0 \circ z) - \frac{f^0 \circ q}{2f^0 \circ s}\frac{\partial}{\partial y}(f^0 \circ s - f^0 \circ z)\right]. \quad \text{(C7)}$$

Expressing $f^0 \circ s, f^0 \circ p, f^0 \circ q$, and $f^0 \circ z$ in terms of usual notations $n, nw_x, nw_y, n(w_x^2 - w_y^2)$, we have

$$\left[\frac{dD^0}{dt}\right]_k = \frac{nD^0}{2} \sum_{i=1}^{N} \frac{\chi[\vec{u}_i]}{f^0(\vec{u}_i)}\left[(1 - \vec{w}^2)\left(\frac{\partial w_x}{\partial x} - \frac{\partial w_y}{\partial y}\right) + 2w_x w_y\left(\frac{\partial w_y}{\partial x} - \frac{\partial w_x}{\partial y}\right)\right]. \quad \text{(C8)}$$

This expression is used in the deduction of the Navier–Stokes equations (5.4a).

APPENDIX D: DETAIL OF THE CALCULATION OF THE NAVIER–STOKES EQUATIONS

We shall calculate the momentum flux $\vec{C}$, which is formally expressed by

$$C \overset{2}{=} \left.\begin{matrix} p \\ q \end{matrix}\right| \circ \int d\vec{r}_2 d\vec{u}_2 \phi(\vec{r}_1 - \vec{r}_2) T_{12}[f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2)]. \quad \text{(D1)}$$

Since collisions occur only between particles separated by a microscopic distance (smaller than 1), we can develop $f^0(\vec{r}_2, \vec{u}_2)$ in (D1) in Taylor series near $f^0(\vec{r}_1, \vec{u}_2)$:

$$f^0(\vec{r}_2, \vec{u}_2) = K_{12} f^0(\vec{r}_1, \vec{u}_2), \quad \text{(D2)}$$

where K_{12} is a translation operator. With the aid of this operator we have

$$T_{12}[f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2)] = T_{12}[f^0(\vec{r}_1, \vec{u}_1) K_{12} f^0(\vec{r}_1, \vec{u}_2)]. \quad \text{(D3)}$$

Performing the scalar products in the space of the one-body distribution functions

$$\int d\vec{u}_2 T_{12}[f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2)] \circ p = -(p + q) \circ \int d\vec{u}_2 T_{12}[f^0(\vec{r}_1, \vec{u}_1) K_{12} f^0(\vec{r}_1, \vec{u}_2)], \quad \text{(D4a)}$$

$$\int d\vec{u}_2 T_{12}[f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2)] \circ q = (p - q) \circ \int d\vec{u}_2 T_{12}[f^0(\vec{r}_1, \vec{u}_1) K_{12} f^0(\vec{r}_1, \vec{u}_2)]. \quad \text{(D4b)}$$

The translation operator K_{12} can be written when a collision occurs; $(\vec{r}_1 - \vec{r}_2)$ is then parallel ($\epsilon = +1$) or antiparallel ($\epsilon = -1$) to the vector $(1, 1)$:

$$K_{12} = 1 + \epsilon \frac{|\vec{r}_1 - \vec{r}_2|}{\sqrt{2}}\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) + \frac{1}{2}\left(\frac{|\vec{r}_1 - \vec{r}_2|}{\sqrt{2}}\right)^2\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 + \cdots. \quad \text{(D5)}$$

We note that the integration domain in the right-hand side of (D1) is symmetric for the interchange of the indices 1 and 2, or equivalently, for the transformation $\epsilon \to -\epsilon$. It is then of no use to calculate the parts of $T_{12} f^0(\vec{r}_1, \vec{u}_1) \times f^0(\vec{r}_2, \vec{u})$ proportional to ϵ. The remaining part reads

$$T_{12}[f^0(\vec{r}_1, \vec{u}_1) f^0(\vec{r}_2, \vec{u}_2)]_R \overset{2}{=} T_{12}\left[f^0(\vec{r}_1, \vec{u}_1) \times \frac{1}{2}\left(\frac{|\vec{r}_1 - \vec{r}_2|}{\sqrt{2}}\right)^2 \times \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 f^0(\vec{r}_1, \vec{u}_2)\right]. \quad \text{(D6)}$$

Then the formulas (B2) enable us to find the expression for $\vec{C}$:

$$C_x \doteq \frac{2}{24}\frac{n^2}{}\left[w_x\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2(w_x^2-w_y^2) - (1+w_x^2-w_y^2)\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2 w_x\right] \tag{D7a}$$

and

$$C_y \doteq \frac{2}{24}\frac{n^2}{}\left[w_y\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2(w_y^2-w_x^2) - (1+w_y^2-w_x^2)\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2 w_y\right] \tag{D7b}$$

---

* Laboratoire associe au Centre National de la Recherche Scientifique.

1 R. Zwanzig, Ann. Rev. Phys. Chem. **16**, 67 (1965).
2 B. J. Alder and T. E. Wainwright, Phys. Rev. A **1**, 18 (1970); M. H. Ernst, E. H. Hange, and J. M. J. Van Leeuwen, Phys. Rev. Letters 25, 1259 (1970).
3 B. J. Alder, D. M. Gass, and T. E. Wainwright, J. Chem. Phys. 53, 3813 (1970).
4 J. C. Maxwell, *Scientific Papers II* (Cambridge U.P., Cambridge, England, 1890), p. 26.
5 J. Broadwell, Phys. Fluids **9**, 1328 (1969).
6 S. Harris, Phys. Fluids 7, 1243 (1964).
7 J. J. Hopfield and J. F. Bastin, Phys. Rev. **168**, 193 (1968).

Algebraic Solution for the Källén-Pauli State Vectors in the $V\theta$ Sector by a Congruence Transformation

C. A. Nelson
(Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803
(Received 9 November 1971; Revised Manuscript Received 7 February 1972)

From the representation of Bolsterli of the $V\theta$ sector physical states in a basis having singular integral equations of the separable type, a congruence transformation carries the representation into the Källén–Pauli Heisenberg field components. This solves the Källén–Pauli singular integral equation algebraically. The resulting state vectors are proven explicitly to be both complete and orthonormal and furnish a new Möller wave matrix.

INTRODUCTION

The S-matrix elements for the $V\theta$ sector have been obtained algebraically by Bolsterli[1] by means of a representation of the physical states in an oblique overcomplete basis. We construct the required congruence transformation to obtain the Heisenberg field components of the Tamm–Dancoff expansion and thereby solve by algebraic techniques the celebrated Källén–Pauli integral equation.[2]

REVIEW: ALGEBRAIC REPRESENTATION OF BOLSTERLI OF THE PHYSICAL STATES IN A RECIPROCAL BASIS

The model has continuum channel $N\theta$ coupled to a discrete "bound state" V with the possible transition $V \rightleftharpoons N\theta$. It is defined by the Hamiltonian

$$\begin{aligned} H &= H_0 + H_I,\\ H_0 &= m^0 V^\dagger V + \int d\mathbf{k}\,\omega(\mathbf{k}) a^\dagger(\mathbf{k})a(\mathbf{k}),\\ H_I &= \frac{1}{(4\pi)^{1/2}}\int d\mathbf{k}\,\frac{f(\omega)}{(2\omega)^{1/2}}[V^\dagger N a(\mathbf{k}) + N^\dagger V a^\dagger(\mathbf{k})], \end{aligned} \tag{1}$$

with[3]

$$\begin{aligned} &\omega(\mathbf{k}) = (\mu^2+\mathbf{k}^2)^{1/2}, \quad \text{where}\\ &[V, V^\dagger] = [N, N^\dagger] = 1,\\ &[a(\mathbf{k}), a^\dagger(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}'). \end{aligned} \tag{2}$$

As Bolsterli has done to obtain the S-matrix elements, we add a (bare) meson to the *physical* $N\theta$ sector states, $|V\rangle = V^\dagger|0\rangle$ and $|N\rangle = N^\dagger|0\rangle$,

$$\begin{aligned} |V\rangle\rangle &= Z^{1/2}|V\rangle + \int d\mathbf{k} F(\mathbf{k}) a^\dagger(\mathbf{k})|N\rangle,\\ |\mathbf{k}\rangle\rangle &= g^*(\mathbf{k})|V\rangle + \int d\mathbf{k}' G(\mathbf{k},\mathbf{k}') a^\dagger(\mathbf{k}')|N\rangle, \end{aligned} \tag{3}$$

where $|V\rangle\rangle$ is the discrete physical V_r state, defined by $\alpha(M)=0$, and $|\mathbf{k}\rangle\rangle \equiv |\mathbf{k}^{\text{in}}\rangle\rangle$ is the "in state" describing the physical $N\theta$ continuum. Our notation is

$$\begin{aligned} \alpha^\pm(\chi) &= \alpha(\chi \pm i\epsilon),\\ \alpha^+(\chi) &= \chi - m^0 - \int d\mathbf{l}\,\frac{f(\nu)^2}{(8\pi\nu)(\chi-\nu+i\epsilon)},\\ Z &= \left(1 + \int d\mathbf{l}\,\frac{f(\nu)^2}{(8\pi\nu)(M-\nu)^2}\right)^{-1},\\ g(\mathbf{k}) &= \frac{f(\omega)}{(8\pi\omega)^{1/2}\alpha^-(\omega)},\\ F(\mathbf{l}) &= \frac{Z^{1/2}f(\nu)}{(8\pi\nu)^{1/2}(M-\nu)}\\ G(\mathbf{k},\mathbf{l}) &= \delta(\mathbf{l}-\mathbf{k}) + \frac{f(\nu)g^*(\mathbf{k})}{(8\pi\nu)^{1/2}(\omega-\nu+i\epsilon)}. \end{aligned} \tag{4}$$

The cutoff function $f(\nu)$ is such that there are no ghosts and $M < \mu$. Thereby the identities

$$\begin{aligned} 0 &= \langle\langle V|a(\mathbf{l})[H-\lambda]|\Phi\rangle\rangle,\\ 0 &= \langle\langle \mathbf{k}|a(\mathbf{l})[H-\lambda]|\Phi\rangle\rangle \end{aligned}$$

yield the separable-like equations

$$\begin{aligned} [\lambda - M - \nu]\varphi_\lambda(\mathbf{l}) &= \frac{f(\nu)}{(8\pi\nu)^{1/2}}\chi_\lambda,\\ [\lambda-\nu-\omega]\varphi_\lambda(\mathbf{k},\mathbf{l}) &= \frac{f(\nu)}{(8\pi\nu)^{1/2}}\chi_\lambda(\mathbf{k}) \end{aligned} \tag{5a}$$

in terms of the coordinates representing the most general physical state, $|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle$, in the reciprocal "$R$ basis,"

$$\varphi_{\mathbf{q}_1,\mathbf{q}_2}(\mathbf{l}) = \langle\langle V|a(\mathbf{l})|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle,$$

Then the formulas (B2) enable us to find the expression for $\vec{C}$:

$$C_x \doteq \frac{2}{24}\frac{n^2}{}\left[w_x\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2(w_x^2-w_y^2) - (1+w_x^2-w_y^2)\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2 w_x\right] \quad \text{(D7a)}$$

and

$$C_y \doteq \frac{2}{24}\frac{n^2}{}\left[w_y\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2(w_y^2-w_x^2) - (1+w_y^2-w_x^2)\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)^2 w_y\right] \quad \text{(D7b)}$$

---

* Laboratoire associe au Centre National de la Recherche Scientifique.

1 R. Zwanzig, Ann. Rev. Phys. Chem. **16**, 67 (1965).
2 B. J. Alder and T. E. Wainwright, Phys. Rev. A **1**, 18 (1970); M. H. Ernst, E. H. Hange, and J. M. J. Van Leeuwen, Phys. Rev. Letters 25, 1259 (1970).
3 B. J. Alder, D. M. Gass, and T. E. Wainwright, J. Chem. Phys. 53, 3813 (1970).
4 J. C. Maxwell, *Scientific Papers II* (Cambridge U.P., Cambridge, England, 1890), p. 26.
5 J. Broadwell, Phys. Fluids **9**, 1328 (1969).
6 S. Harris, Phys. Fluids **7**, 1243 (1964).
7 J. J. Hopfield and J. F. Bastin, Phys. Rev. **168**, 193 (1968).

Algebraic Solution for the Källén-Pauli State Vectors in the $V\theta$ Sector by a Congruence Transformation

C. A. Nelson

(Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803

(Received 9 November 1971; Revised Manuscript Received 7 February 1972)

From the representation of Bolsterli of the $V\theta$ sector physical states in a basis having singular integral equations of the separable type, a congruence transformation carries the representation into the Källén–Pauli Heisenberg field components. This solves the Källén–Pauli singular integral equation algebraically. The resulting state vectors are proven explicitly to be both complete and orthonormal and furnish a new Möller wave matrix.

INTRODUCTION

The S-matrix elements for the $V\theta$ sector have been obtained algebraically by Bolsterli[1] by means of a representation of the physical states in an oblique overcomplete basis. We construct the required congruence transformation to obtain the Heisenberg field components of the Tamm–Dancoff expansion and thereby solve by algebraic techniques the celebrated Källén–Pauli integral equation.[2]

REVIEW: ALGEBRAIC REPRESENTATION OF BOLSTERLI OF THE PHYSICAL STATES IN A RECIPROCAL BASIS

The model has continuum channel $N\theta$ coupled to a discrete "bound state" V with the possible transition $V \rightleftharpoons N\theta$. It is defined by the Hamiltonian

$$\begin{aligned} H &= H_0 + H_I,\\ H_0 &= m^0 V^\dagger V + \int d\mathbf{k}\,\omega(\mathbf{k}) a^\dagger(\mathbf{k}) a(\mathbf{k}),\\ H_I &= \frac{1}{(4\pi)^{1/2}}\int d\mathbf{k}\,\frac{f(\omega)}{(2\omega)^{1/2}}\left[V^\dagger N a(\mathbf{k}) + N^\dagger V a^\dagger(\mathbf{k})\right], \end{aligned} \quad (1)$$

with[3]

$$\begin{aligned} &\omega(\mathbf{k}) = (\mu^2+\mathbf{k}^2)^{1/2}, \quad \text{where}\\ &[V, V^\dagger] = [N, N^\dagger] = 1,\\ &[a(\mathbf{k}), a^\dagger(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}'). \end{aligned} \quad (2)$$

As Bolsterli has done to obtain the S-matrix elements, we add a (bare) meson to the *physical* $N\theta$ sector states, $|V\rangle = V^\dagger|0\rangle$ and $|N\rangle = N^\dagger|0\rangle$,

$$\begin{aligned} |V\rangle\rangle &= Z^{1/2}|V\rangle + \int d\mathbf{k} F(\mathbf{k}) a^\dagger(\mathbf{k})|N\rangle,\\ |\mathbf{k}\rangle\rangle &= g^*(\mathbf{k})|V\rangle + \int d\mathbf{k}' G(\mathbf{k},\mathbf{k}') a^\dagger(\mathbf{k}')|N\rangle, \end{aligned} \quad (3)$$

where $|V\rangle\rangle$ is the discrete physical V_r state, defined by $\alpha(M) = 0$, and $|\mathbf{k}\rangle\rangle \equiv |\mathbf{k}^{\text{in}}\rangle\rangle$ is the "in state" describing the physical $N\theta$ continuum. Our notation is

$$\begin{aligned} \alpha^\pm(\chi) &= \alpha(\chi \pm i\epsilon),\\ \alpha^+(\chi) &= \chi - m^0 - \int d\mathbf{l}\,\frac{f(\nu)^2}{(8\pi\nu)(\chi-\nu+i\epsilon)},\\ Z &= \left(1+\int d\mathbf{l}\,\frac{f(\nu)^2}{(8\pi\nu)(M-\nu)^2}\right)^{-1},\\ g(\mathbf{k}) &= \frac{f(\omega)}{(8\pi\omega)^{1/2}\alpha^-(\omega)},\\ F(\mathbf{l}) &= \frac{Z^{1/2} f(\nu)}{(8\pi\nu)^{1/2}(M-\nu)}\\ G(\mathbf{k},\mathbf{l}) &= \delta(\mathbf{l}-\mathbf{k}) + \frac{f(\nu) g^*(\mathbf{k})}{(8\pi\nu)^{1/2}(\omega-\nu+i\epsilon)}. \end{aligned} \quad (4)$$

The cutoff function $f(\nu)$ is such that there are no ghosts and $M < \mu$. Thereby the identities

$$\begin{aligned} 0 &= \langle\langle V|a(\mathbf{l})[H-\lambda]|\Phi\rangle\rangle,\\ 0 &= \langle\langle\mathbf{k}|a(\mathbf{l})[H-\lambda]|\Phi\rangle\rangle \end{aligned}$$

yield the separable-like equations

$$\begin{aligned} [\lambda - M - \nu]\varphi_\lambda(\mathbf{l}) &= \frac{f(\nu)}{(8\pi\nu)^{1/2}}\chi_\lambda,\\ [\lambda-\nu-\omega]\varphi_\lambda(\mathbf{k},\mathbf{l}) &= \frac{f(\nu)}{(8\pi\nu)^{1/2}}\chi_\lambda(\mathbf{k}) \end{aligned} \quad \text{(5a)}$$

in terms of the coordinates representing the most general physical state, $|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle$, in the reciprocal "$R$ basis,"

$$\varphi_{\mathbf{q}_1,\mathbf{q}_2}(\mathbf{l}) = \langle\langle V|a(\mathbf{l})|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle,$$

$$\varphi_{\mathbf{q}_1,\mathbf{q}_2}(\mathbf{k},\mathbf{l}) = \langle\langle \mathbf{k}|a(\mathbf{l})|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle,$$
$$\chi_{\mathbf{q}_1,\mathbf{q}_2} = \langle\langle V|N^{\dagger}V|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle,$$
$$\chi_{\mathbf{q}_1,\mathbf{q}_2}(\mathbf{k}) = \langle\langle \mathbf{k}|N^{\dagger}V|\Phi_{\mathbf{q}_1,\mathbf{q}_2}\rangle\rangle. \tag{6}$$

The "R basis" is given by $a^{\dagger}(\mathbf{l})$ and $V^{\dagger}N$ acting on $|V\rangle\rangle$ and $|\mathbf{k}\rangle\rangle$, and it is both an oblique and overcomplete basis. (The subscripts "$\mathbf{q}^1, \mathbf{q}^2$" will be denoted by λ or suppressed.) To close the system of equations,

$$0 = \langle\langle V|N^{\dagger}V[H-\lambda]|\Phi\rangle\rangle,$$
$$0 = \langle\langle \mathbf{k}|N^{\dagger}V[H-\lambda]|\Phi\rangle\rangle$$

give

$$[\lambda - M - m^0]\chi_\lambda = (1-2Z)\int d\mathbf{k}\,\frac{f(\omega)}{(8\pi\omega)^{1/2}}\varphi_\lambda(\mathbf{k}) - 2Z^{1/2}\iint d\mathbf{k}d\mathbf{s}\,\frac{f(\omega)}{(8\pi\omega)^{1/2}}g^*(\mathbf{s})\varphi_\lambda(\mathbf{s},\mathbf{k}),$$

$$[\lambda-\omega-m^0]\chi_\lambda(\mathbf{k}) = \int d\mathbf{t}\,\frac{f(\tau)}{(8\pi\tau)^{1/2}}\varphi_\lambda(\mathbf{k},\mathbf{t}) - 2Z^{1/2}g(\mathbf{k})\int d\mathbf{t}\,\frac{f(\tau)}{(8\pi\tau)^{1/2}}\varphi_\lambda(\mathbf{t}) - 2g(\mathbf{k})\iint d\mathbf{t}d\mathbf{s} \times \frac{f(\tau)}{(8\pi\tau)^{1/2}}g^*(\mathbf{s})\varphi_\lambda(\mathbf{s},\mathbf{t}). \tag{5b}$$

As for the $N\theta$ sector, it is straightforward to solve Eqs. (5). In a columnar notation for the components

$$\Phi_\lambda(\mathbf{k},\mathbf{l}) = \begin{bmatrix}\chi_\lambda\\ \chi_\lambda(\mathbf{k})\\ \varphi_\lambda(\mathbf{l})\\ \varphi_\lambda(\mathbf{k},\mathbf{l})\end{bmatrix},$$

the $V_r\theta$ bound state is represented by $(\Lambda < M+\mu)$

$$\Phi_\Lambda(\mathbf{k},\mathbf{l}) = c\begin{bmatrix}1\\ \bar{\chi}_\Lambda(\mathbf{k})\\ \dfrac{f(\nu)}{(8\pi\nu)^{1/2}(\Lambda-M-\nu)}\cdot 1\\ \dfrac{f(\nu)}{(8\pi\nu)^{1/2}(\Lambda-\omega-\nu)}\cdot\bar{\chi}_\Lambda(\mathbf{k})\end{bmatrix} \tag{6'a}$$

with $c = [2\gamma(\Lambda)/\eta'(\Lambda)]^{1/2}$ and $\bar{\chi}_\Lambda(\mathbf{k}) = g(\mathbf{k})\alpha(\Lambda-M)/Z^{1/2}\alpha(\Lambda-\omega)$. Prime denotes derivative on $\eta'(\Lambda)$. The $V_r\theta$ scattering state by $(\mu \le \xi = \lambda - M < \infty)$

$$\Phi_\xi(\mathbf{k},\mathbf{l}) = \begin{bmatrix}\chi_\xi\\ \chi_\xi(\mathbf{k})\\ \delta(\mathbf{l}-\mathbf{q}) + \dfrac{f(\nu)}{(8\pi\nu)^{1/2}(\xi-\nu+i\epsilon)}\chi_\xi\\ \dfrac{f(\nu)}{(8\pi\nu)^{1/2}(\xi+M-\nu-\omega+i\epsilon)}\cdot\chi_\xi(\mathbf{k})\end{bmatrix} \tag{6'b}$$

with

$$\chi_\xi = g^*(\mathbf{q})\left(1-\frac{2Z}{\eta^+(\xi+M)}\right),$$

$$\chi_\xi(\mathbf{k}) = Z^{1/2}\delta(\mathbf{k}-\mathbf{q}) - \frac{2Z^{1/2}g(\mathbf{k})f(\xi)}{(8\pi\xi)^{1/2}\eta^+(\xi+M)\alpha^+(\xi+M-\omega)},$$

where

$$\eta(\lambda) = Z + \gamma(\lambda)\alpha(\lambda-M), \qquad \gamma(\lambda) = \int d\mathbf{k}\,\frac{|g(\mathbf{k})|^2}{\alpha(\lambda-\omega)} \tag{7}$$

and the $N\theta\theta$ scattering state for $\lambda = \xi_1+\xi_2$ by $(\mu < \xi_1, \xi_2 < \infty)$

$$\Phi_{\xi_1,\xi_2}(\mathbf{k},\mathbf{l}) = \begin{bmatrix}\chi_{\xi_1,\xi_2}\\ \chi_{\xi_1,\xi_2}(\mathbf{k})\\ \dfrac{f(\nu)}{(8\pi\nu)^{1/2}(\xi_1+\xi_2-M-\nu+i\epsilon)}\cdot\chi_{\xi_1,\xi_2}\\ \left[\delta(\mathbf{q}_1-\mathbf{l})\delta(\mathbf{q}_2-\mathbf{k}) + \delta(\mathbf{q}_1-\mathbf{k})\delta(\mathbf{q}_2-\mathbf{l}) + \dfrac{f(\nu)}{(8\pi\nu)^{1/2}(\xi_1+\xi_2-\omega-\nu+i\epsilon)}\cdot\chi_{\xi_1,\xi_2}(\mathbf{k})\right]\end{bmatrix} \tag{6'c}$$

with

$$\chi_{\xi_1,\xi_2} = -2Z^{1/2}g^*(\mathbf{q}_1)g^*(\mathbf{q}_2)/\eta^+(\xi_1+\xi_2),$$

$$\chi_{\xi_1,\xi_2}(\mathbf{k}) = g^*(\mathbf{q}_1)\delta(\mathbf{q}_2-\mathbf{k}) + g^*(\mathbf{q}_2)\delta(\mathbf{q}_1-\mathbf{k}) - \frac{2g^*(\mathbf{q}_1)g^*(\mathbf{q}_2)\alpha^+(\xi_1+\xi_2-M)g(\mathbf{k})}{\eta^+(\xi_1+\xi_2)\alpha^+(\xi_1+\xi_2-\omega}.$$

ALGEBRAIC CONGRUENCE TRANSFORMATION FROM THE "R BASIS" TO THE "KÄLLÉN–PAULI BASIS"

Our objective is to obtain the Heisenberg fields describing the physical states of the $V\theta$ sector in terms of the standard orthonormal basis of free field eigenstates. Hence, we must construct the congruence transformation T which takes the metric matrix M in the "R basis" into the diagonal metric matrix, $M' = T^{\dagger}MT$, of the "KP basis."

The oblique overcomplete "R basis" consists of

$$|V) \equiv V^{\dagger}N|V\rangle\rangle = \int d\mathbf{k}'F(\mathbf{k}')a^{\dagger}(\mathbf{k}')|V\rangle,$$
$$|\mathbf{k}) \equiv V^{\dagger}N|\mathbf{k}\rangle\rangle = \int d\mathbf{k}'G(\mathbf{k},\mathbf{k}')a^{\dagger}(\mathbf{k}')|V\rangle,$$
$$|V,\mathbf{l}) \equiv a^{\dagger}(\mathbf{l})|V\rangle\rangle,$$
$$|\mathbf{k},\mathbf{l}) \equiv a^{\dagger}(\mathbf{l})|\mathbf{k}\rangle\rangle.$$

Hence, the metric matrix M of scalar products is

$$M = \begin{bmatrix}(V|V) & (V|\mathbf{k}') & (V|V,\mathbf{l}') & (V|\mathbf{k}',\mathbf{l}')\\ (\mathbf{k}|V) & (\mathbf{k}|\mathbf{k}') & (\mathbf{k}|V,\mathbf{l}') & (\mathbf{k}|\mathbf{k}',\mathbf{l}')\\ (V,\mathbf{l}|V) & (V,\mathbf{l}|\mathbf{k}') & (V,\mathbf{l}|V,\mathbf{l}') & (V,\mathbf{l}|\mathbf{k}',\mathbf{l}')\\ (\mathbf{k},\mathbf{l}|V) & (\mathbf{k},\mathbf{l}|\mathbf{k}') & (\mathbf{k},\mathbf{l}|V,\mathbf{l}') & (\mathbf{k},\mathbf{l}|\mathbf{k}',\mathbf{l}')\end{bmatrix}.$$

Using the equations

$$Z + \int d\mathbf{l}' \,|F(\mathbf{l}')|^2 = 1,$$
$$Z^{1/2}g(\mathbf{s}) + \int d\mathbf{l}' F(\mathbf{l}')G^*(\mathbf{s},\mathbf{l}') = 0,$$
$$g^*(\mathbf{l})g(\mathbf{s}) + \int d\mathbf{l}' G(\mathbf{l},\mathbf{l}')G^*(\mathbf{s},\mathbf{l}') = \delta(\mathbf{l}-\mathbf{s}),$$
$$Z + \int d\mathbf{l}' \,|g(\mathbf{l}')|^2 = 1,$$
$$Z^{1/2}F(\mathbf{s}) + \int d\mathbf{l}' g(\mathbf{l}')G(\mathbf{l}',\mathbf{s}) = 0,$$
$$F^*(\mathbf{l})F(\mathbf{s}) + \int d\mathbf{l}' G^*(\mathbf{l}',\mathbf{l})G(\mathbf{l}',\mathbf{s}) = \delta(\mathbf{l}-\mathbf{s}), \tag{8}$$

which follow from the unitarity of the generalized Möller matrix in the $N\theta$ sector, we obtain

$$M = \begin{bmatrix} 1-Z & -Z^{1/2}g^*(\mathbf{k}') & Z^{1/2}F^*(\mathbf{l}') & g^*(\mathbf{k}')F^*(\mathbf{l}') \\ -Z^{1/2}g(\mathbf{k}) & [\delta(\mathbf{k}-\mathbf{k}') - g(\mathbf{k})g^*(\mathbf{k}')] & Z^{1/2}G^*(\mathbf{k},\mathbf{l}') & g^*(\mathbf{k}')G^*(\mathbf{k},\mathbf{l}') \\ Z^{1/2}F(\mathbf{l}) & Z^{1/2}G(\mathbf{k}',\mathbf{l}) & [\delta(\mathbf{l}-\mathbf{l}') + F(\mathbf{l})F^*(\mathbf{l}')] & G(\mathbf{k}',\mathbf{l})F^*(\mathbf{l}') \\ g(\mathbf{k})F(\mathbf{l}) & g(\mathbf{k})G(\mathbf{k}',\mathbf{l}) & F(\mathbf{l})G^*(\mathbf{k},\mathbf{l}') & [\delta(\mathbf{l}-\mathbf{l}')\delta(\mathbf{k}-\mathbf{k}') + G^*(\mathbf{k},\mathbf{l}')G(\mathbf{k}',\mathbf{l})] \end{bmatrix}. \tag{9}$$

Notice that M is Hermitian so that the congruence transformation T is the modal matrix composed of the eigenvectors of M. This fact and inspection of M indicate that it is simplest to diagonalize M by a series of congruence transformations, each of which successively further reduces M until the diagonal matrix M' is obtained.

We begin by diagonalizing with a congruence transformation the submatrix m, given by

$$m = \begin{bmatrix} \delta(\mathbf{l}-\mathbf{l}') + F(\mathbf{l})F(\mathbf{l}') & G(\mathbf{k}',\mathbf{l})F^*(\mathbf{l}') \\ F(\mathbf{l})G^*(\mathbf{k},\mathbf{l}') & \delta(\mathbf{l}-\mathbf{l}')\delta(\mathbf{k}-\mathbf{k}') + G^*(\mathbf{k},\mathbf{l}')G(\mathbf{k}',\mathbf{l}) \end{bmatrix} \tag{10}$$

and then employ its eigenvectors to reduce M: Let

$$X_t = \begin{bmatrix} a_t(\mathbf{l}') \\ b_t(\mathbf{l}'\mathbf{k}') \end{bmatrix}$$

be an eigenvector so that the eigenvalue equations are

$$a_t(\mathbf{l}) + F(\mathbf{l})\int d\mathbf{l}' F^*(\mathbf{l}')a_t(\mathbf{l}') + \int d\mathbf{k}' G(\mathbf{k}',\mathbf{l})\int d\mathbf{l}' F^*(\mathbf{l}')b_t(\mathbf{l}',\mathbf{k}') = \lambda_t a_t(\mathbf{l}),$$
$$F(\mathbf{l})\int d\mathbf{l}' G^*(\mathbf{k},\mathbf{l}')a_t(\mathbf{l}') + b_t(\mathbf{l},\mathbf{k}) + \int d\mathbf{k}' G(\mathbf{k}',\mathbf{l})\int d\mathbf{l}' G^*(\mathbf{k},\mathbf{l}')b_t(\mathbf{l}',\mathbf{k}') = \lambda_t b_t(\mathbf{l},\mathbf{k}). \tag{11}$$

Using Eqs. (8) by algebraic iteration, we determine the eigenvalues to be $\lambda_t = 1, 2,$ and 0 (the last a consequence of redundancy in the m subspace). For $\lambda_t = 1$, Eqs. (11) require that the projection of

$$F(\mathbf{l})a_t(\mathbf{l}') + \int d\mathbf{k}' G(\mathbf{k}'\mathbf{l})b_t(\mathbf{l}',\mathbf{k}'),$$

with both $F^*(\mathbf{l})$ and $G(\mathbf{k},\mathbf{l})$ vanish. Hence, by comparing Eqs. (8), for $\lambda_{t_1} = 1$

$$X_{t_1} = \delta(\mathbf{l}-\mathbf{l}_t)\begin{bmatrix} Z^{1/2} \\ g(\mathbf{k}) \end{bmatrix}. \tag{12}$$

Similarly, for $\lambda_{t_2} = 2$ we find

$$X_{t_2} = \tfrac{1}{2}\begin{bmatrix} \delta(\mathbf{l}-\mathbf{l}_t)F^*(\mathbf{k}_t) + \delta(\mathbf{l}-\mathbf{k}_t)F^*(\mathbf{l}_t) \\ \delta(\mathbf{l}-\mathbf{l}_t)G^*(\mathbf{k},\mathbf{k}_t) + \delta(\mathbf{l}-\mathbf{k}_t)G^*(\mathbf{k},\mathbf{l}_t) \end{bmatrix}. \tag{13}$$

Next we act on M with a matrix A, constructed out of these eigenvectors,

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \delta(\mathbf{k}'-\mathbf{t}') & 0 & 0 \\ 0 & 0 & Z^{1/2}\delta(\mathbf{l}'-\mathbf{s}') & \tfrac{1}{2}[\delta(\mathbf{l}'-\mathbf{s}')F^*(\mathbf{t}') + \delta(\mathbf{l}'-\mathbf{t}')F^*(\mathbf{s}')] \\ 0 & 0 & g(\mathbf{k}')\delta(\mathbf{l}'-\mathbf{s}') & \tfrac{1}{2}[\delta(\mathbf{l}'-\mathbf{s}')G^*(\mathbf{k}',\mathbf{t}') + \delta(\mathbf{l}'-\mathbf{t}')G^*(\mathbf{k}',\mathbf{s}')] \end{bmatrix}, \tag{14}$$

and obtain the reduced matrix

$$A^+MA = \begin{bmatrix} 1-Z & -Z^{1/2}g^*(\mathbf{t}') & F^*(\mathbf{s}') & 0 \\ -Z^{1/2}g(\mathbf{t}) & [\delta(\mathbf{t}-\mathbf{t}') - g(\mathbf{t})g^*(\mathbf{t}')] & G^*(\mathbf{t}',\mathbf{s}') & 0 \\ F(\mathbf{s}) & G(\mathbf{t}',\mathbf{s}) & \delta(\mathbf{s}-\mathbf{s}') & 0 \\ 0 & 0 & 0 & \delta(\mathbf{s}-\mathbf{s}')\delta(\mathbf{t}-\mathbf{t}' + \delta(\mathbf{s}-\mathbf{t}')\delta(\mathbf{t}-\mathbf{s}') \end{bmatrix}. \tag{15}$$

Repeating this procedure for the remaining submatrix, we find that

$$B = \tfrac{1}{2}\begin{bmatrix} \sqrt{2} & -F^*(\mathbf{t}) & F^*(\mathbf{s}) & 0 \\ 0 & G^*(\mathbf{t}',\mathbf{t}) & G^*(\mathbf{t}',\mathbf{s}) & 0 \\ 0 & \delta(\mathbf{s}'-\mathbf{t}) & \delta(\mathbf{s}'-\mathbf{s}) & 0 \\ 0 & 0 & 0 & \tfrac{1}{2}[\delta(\mathbf{s}'-\mathbf{s})\delta(\mathbf{t}'-\mathbf{t}) + \delta(\mathbf{s}'-\mathbf{t})\delta(\mathbf{t}'-\mathbf{s})] \end{bmatrix} \tag{16}$$

and

$$C = \tfrac{1}{2}\begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \delta(\mathbf{t}-\mathbf{q}') & 0 & 0 \\ -F(\mathbf{s}) & 0 & \delta(\mathbf{s}-\mathbf{m}') & 0 \\ 0 & 0 & 0 & \tfrac{1}{2}[\delta(\mathbf{s}-\mathbf{g}')\delta(\mathbf{t}-\mathbf{m}') + \delta(\mathbf{s}-\mathbf{m}')\delta(\mathbf{t}-\mathbf{g}')] \end{bmatrix} \tag{16}$$

are needed to obtain the diagonal matrix $M' = T^+ M\, T$:

$$M' = \operatorname{diag}(0, 0, \delta(\mathbf{m}-\mathbf{m}'), \delta(\mathbf{m}-\mathbf{m}')\delta(\mathbf{g}-\mathbf{g}') + \delta(\mathbf{m}-\mathbf{g}')\delta(\mathbf{g}-\mathbf{m}')), \tag{17}$$

where $T = A\,BC$ or

$$T = \tfrac{1}{2}\begin{bmatrix} 1+Z & -F^*(\mathbf{g}') & F^*(\mathbf{m}') & 0 \\ -Z^{1/2}g(\mathbf{k}') & G^*(\mathbf{k}',\mathbf{g}') & G^*(\mathbf{k}',\mathbf{m}') & 0 \\ -Z^{1/2}F(\mathbf{l}') & Z^{1/2}\delta(\mathbf{l}'-\mathbf{g}') & Z^{1/2}\delta(\mathbf{l}'-\mathbf{m}') & [\delta(\mathbf{l}'-\mathbf{g}')F^*(\mathbf{m}') + \delta(\mathbf{l}'-\mathbf{m}')F^*(\mathbf{g}')] \\ -g(\mathbf{k})F(\mathbf{l}') & g(\mathbf{k}')\delta(\mathbf{l}'-\mathbf{g}') & g(\mathbf{k}')\delta(\mathbf{l}'-\mathbf{m}') & [\delta(\mathbf{l}'-\mathbf{g}')G^*(\mathbf{k}',\mathbf{m}') + \delta(\mathbf{l}'-\mathbf{m}')G^*(\mathbf{k}',\mathbf{g}')] \end{bmatrix} \tag{18}$$

With

$$|e_j) \equiv \begin{bmatrix} |V) \\ |\mathbf{k}') \\ |V,\mathbf{l}') \\ |\mathbf{k}',\mathbf{l}') \end{bmatrix},$$

the "R basis" is transformed to (T^T is transpose of T)

$$|e_i'\rangle = (T^T)_{ij}\,|e_j) = \begin{bmatrix} 0 \\ 0 \\ a^+(\mathbf{m}')\,|V\rangle \\ a^+(\mathbf{m}')a^+(\mathbf{g}')\,|N\rangle \end{bmatrix}$$

that is, the "KP basis"!

Thus, the reciprocal components of the "R basis," given by $\Phi_{\mathbf{q}_1,\mathbf{q}_2}(\mathbf{k}',\mathbf{l}')$, are transformed into

$$\begin{bmatrix} 0 \\ 0 \\ \psi(\mathbf{m}) \\ \psi(\mathbf{g},\mathbf{m}) \end{bmatrix} = T^T \begin{bmatrix} \chi \\ \chi(\mathbf{k}') \\ \varphi(\mathbf{l}') \\ \varphi(\mathbf{k}',\mathbf{l}') \end{bmatrix},$$

where $\psi(\mathbf{m})$ and $\psi(\mathbf{g},\mathbf{m})$ are desired Källén–Pauli components

$$|\Phi\rangle\rangle = \int d\mathbf{l}\,\psi(\mathbf{l})a^+(\mathbf{l})\,|V\rangle + \tfrac{1}{2}\iint d\mathbf{k}_1 d\mathbf{k}_2\,\psi(\mathbf{k}_1,\mathbf{k}_2)a^+(\mathbf{k}_1)a^+(\mathbf{k}_2)\,|N\rangle. \tag{19}$$

That is, we have obtained the very simple result

$$\psi(\mathbf{l}) = Z^{1/2}\varphi(\mathbf{l}) + \int d\mathbf{k}\,g^*(\mathbf{k})\varphi(\mathbf{k},\mathbf{l}), \tag{20a}$$

$$\psi(\mathbf{k}_1,\mathbf{k}_2) = \tfrac{1}{2}[\varphi(\mathbf{k}_1)F(\mathbf{k}_2) + \int d\mathbf{k}\,G(\mathbf{k},\mathbf{k}_1)\varphi(\mathbf{k}_1,\mathbf{k}_2) + (\mathbf{k}_1 \leftrightarrow \mathbf{k}_2)], \tag{20b}$$

and by linear dependence the relations

$$\begin{aligned} &Z^{1/2}\chi + \int d\mathbf{k}\,g^*(\mathbf{k})\chi(\mathbf{k}) = 0, \\ &F(\mathbf{l})\chi + \int d\mathbf{k}\,G(\mathbf{k},\mathbf{l})\chi(\mathbf{k}) - Z^{1/2}\varphi(\mathbf{l}) - \int d\mathbf{k}\,g^*(\mathbf{k})\varphi(\mathbf{k},\mathbf{l}) = 0 \end{aligned} \tag{21}$$

as a precise statement of the redundancy of the "R basis." It is straightforward to verify directly that Eqs. (20) satisfy the Källén–Pauli eigenequations

$$\begin{aligned} [\lambda - \sigma - \tau]\psi(\mathbf{s},\mathbf{t}) &= \frac{f(\sigma)}{(8\pi\sigma)^{1/2}}\,\psi(\mathbf{t}) + \frac{f(\tau)}{(8\pi\tau)^{1/2}}\,\psi(\mathbf{s}), \\ [\lambda - m^0 - \sigma]\psi(\mathbf{s}) &= \frac{1}{(4\pi)^{1/2}}\int d\mathbf{l}\,\frac{f(\nu)}{(2\nu)^{1/2}}\,\psi(\mathbf{l},\mathbf{s}). \end{aligned} \tag{22}$$

Simply substitute and use the "R basis" eigenequations, the definition of $F(\mathbf{l})$, and Eqs. (21).

The simplified Källén–Pauli state vectors are

(i) V,θ bound state defined by $\eta(\Lambda) = 0$ ($\Lambda < M + \mu$):

$$\psi_\Lambda(\mathbf{l})\ \frac{c\alpha(\Lambda - M)}{Z^{1/2}}\ \frac{f(\nu)}{(8\pi\nu)^{1/2}(\nu - \Lambda + M)} \times \int d\mathbf{k}\,\frac{|g(\mathbf{k})|^2(\omega - M)}{(\omega - \Lambda + \nu)\alpha(\Lambda - \omega} \tag{23a}$$

with

$$c = [2\gamma(\Lambda)/\eta'(\Lambda)]^{1/2}, \quad \psi_\Lambda(\mathbf{k}_1,\mathbf{k}_2) = \frac{[\psi_\Lambda(\mathbf{k}_1)f(\omega_2)/(8\pi\omega_2)^{1/2} + \psi_\Lambda(\mathbf{k}_2)f(\omega_1)/(8\pi\omega_1)^{1/2}]}{\Lambda - \omega_1 - \omega_2}; \tag{23b}$$

(ii) V,θ scattering state ($\mu \le \xi = \lambda - M < \infty$):

$$\psi_\xi(\mathbf{l}) = Z^{1/2}\delta(\mathbf{l} - \mathbf{q}) - \frac{Z^{1/2}f(\nu)f(\xi)}{(8\pi\nu)^{1/2}(8\pi\xi)^{1/2}(\xi - \nu + i\epsilon)} \times\left(\frac{M - \xi}{\alpha^+(\xi)(M - \nu + i\epsilon)} + \frac{2}{\eta^+(\xi + M)} \times \int d\mathbf{k}\,\frac{|g(\mathbf{k})|^2(\omega - M)}{(\xi - \nu - \omega + M + i\epsilon)\alpha^+(\xi + M - \omega)}\right), \tag{24a}$$

$$\psi_\xi(\mathbf{k}_1,\mathbf{k}_2) = F(\mathbf{k}_2)\delta(\mathbf{k}_1 - \mathbf{q}) - \frac{Z^{1/2}f(\omega_1)f(\omega_2)f(\xi)}{(8\pi\omega_1)^{1/2}(8\pi\omega_2)^{1/2}(8\pi\xi)^{1/2}\eta^+(\xi + M)}$$

$$\times\Big(\frac{1}{(\xi+M-\omega_1-\omega_2+i\epsilon)\alpha^+(\xi+M-\omega_1)\alpha^-(\omega_1)}$$
$$+\frac{1}{(\xi-\omega_2+i\epsilon)(M-\omega_1)}$$
$$\times\int d\mathbf{k}\,\frac{|g(\mathbf{k})|^2(M-\omega)(\omega-\omega_1+\omega_2-\xi)}{(\omega-\omega_1+i\epsilon)(\xi+M-\omega_2-\omega+i\epsilon)}\Big)$$
$$+(\mathbf{k}_1\leftrightarrow\mathbf{k}_2); \tag{24b}$$

(iii) $N\theta\theta$ scattering state ($\lambda=\xi_1+\xi_2, \mu\le\xi_1,\xi_2<\infty$):

$$\psi_{\xi_1,\xi_2}(\mathbf{l})=g^*(\mathbf{q}_1)\delta(\mathbf{q}_2-\mathbf{l})+\frac{f(\nu)g^*(\mathbf{q}_1)g^*(\mathbf{q}_2)}{(8\pi\nu)^{1/2}(\xi_1-\nu+i\epsilon)}$$
$$-\frac{f(\nu)g^*(q_1)g^*(q_2)}{(8\pi\nu)^{1/2}(\xi_1+\xi_2-M-\nu+i\epsilon)}$$
$$\times\Big(1+\frac{\alpha^+(\xi_1+\xi_2-M)}{\eta^+(\xi_1+\xi_2)}$$
$$\times\int dk\,|g(k)|^2(\omega-M)/$$
$$\times\begin{bmatrix}(\xi_1+\xi_2-\omega-\nu+i\epsilon)\\ X\alpha^+(\xi_1+\xi_2-\omega)\end{bmatrix}\Big)+(\mathbf{q}_1\leftrightarrow\mathbf{q}_2), \tag{25a}$$

$$\psi_{\xi_1,\xi_2}(\mathbf{k}_1,\mathbf{k}_2)$$
$$=\tfrac{1}{2}[\delta(\mathbf{q}_1-\mathbf{k}_1)\delta(\mathbf{q}_2-\mathbf{k}_2)+\delta(\mathbf{q}_1-\mathbf{k}_2)\xi(\mathbf{q}_2-\mathbf{k}_1)]$$
$$+\Big(\frac{f(\omega_2)g^*(\mathbf{q}_1)\delta(\mathbf{q}_2-\mathbf{k}_1)}{(8\pi\omega_2)^{1/2}(\xi_1-\omega_2+i\epsilon)}+(\mathbf{q}_2\leftrightarrow\mathbf{q}_1)\Big)$$
$$+\frac{f(\omega_1)f(\omega_2)g^*(\mathbf{q}_1)g^*(\mathbf{q}_2)}{(8\pi\omega_1)^{1/2}(8\pi\omega_2)^{1/2}}$$
$$\times\Omega_{\xi_1,\xi_2}(\mathbf{k}_1,\mathbf{k}_2)+(\mathbf{k}_1\leftrightarrow\mathbf{k}_2) \tag{25b}$$

with

$$\Omega_{\xi_1,\xi_2}(\mathbf{k}_1,\mathbf{k}_2)=\frac{1}{(\xi_1-\omega_2+i\epsilon)(\xi_2-\omega_1+i\epsilon)}$$
$$-\alpha^+(\xi_1+\xi_2-M)/$$
$$\times\begin{bmatrix}(\xi_1+\xi_2-\omega_1-\omega_2+i\epsilon)\alpha^+(\xi_1+\xi_2-\omega_1)\\ \times\,\alpha^-(\omega_1)\eta^+(\xi_1+\xi_2)\end{bmatrix}$$
$$-\frac{1}{(M-\omega_2)(\xi_1+\xi_2-M-\omega_1+i\epsilon)}$$
$$\times\Big(1-\frac{\alpha^+(\xi_1+\xi_2-M)}{\eta^+(\xi_1+\xi_2)}$$
$$\times\int d\mathbf{k}\frac{|g(k)|^2(M-\omega)(\xi_1+\xi_2-M-\omega+\omega_1-\omega_2)}{(\omega-\omega_2+i\epsilon)(\xi_1+\xi_2-\omega_1-\omega+i\epsilon)}\Big).$$

The $\psi(\mathbf{l})$ followed from Eq. (20a) by using $\eta(\lambda)=Z+\gamma(\lambda)\alpha(\lambda-M)$ and the spectral representation for $\gamma(\lambda)$. The $\psi(\mathbf{k}_1,\mathbf{k}_2)$ followed from Eq. (20b) in the same way with the $(\mathbf{k}_1\leftrightarrow\mathbf{k}_2)$ symmetry used to combine terms. This form for $\psi(\mathbf{l})$ and $\psi(\mathbf{k}_1,\mathbf{k}_2)$ has the advantage of displaying directly the singularities whose residues are the S matrix elements for the $V\theta$ sector. Pagnamenta's form[4] can be obtained by operations similar to the above and by using the representation

$$\frac{1}{\alpha(\lambda)}=\frac{Z}{\lambda-M}+\int d\mathbf{k}\,\frac{|g(\mathbf{k})|^2}{\lambda-\omega}$$

(The results agree except for the factor of $\frac{1}{2}$ on Eq. (60) in Ref. 4).

APPENDIX

To explicity prove that the Källén–Pauli Heisenberg field components furnish a new Möller wave matrix it is simplest to work in the "R basis." In the usual manner by contour integration the state vectors are shown to satisfy the orthonormality relation.[1] The dual relationship of completeness for the physical states is specified by

$$1=|\Phi_\Lambda\rangle\rangle\langle\langle\Phi_\Lambda|+\int d\mathbf{q}\,|\Phi_\xi\rangle\rangle\langle\langle\Phi_\xi|+\tfrac{1}{2}\iint d\mathbf{q}_1d\mathbf{q}_2\,|\Phi_{\xi_1,\xi_2}\rangle\rangle\langle\langle\Phi_{\xi_1,\xi_2}|.$$

Hence, by taking scalar products of the "R basis," the completeness statement for the physical states is

$$M=\sum_\lambda\begin{pmatrix}\chi\\ \chi(\mathbf{k})\\ \varphi(\mathbf{l})\\ \varphi(\mathbf{k},\mathbf{l})\end{pmatrix}(\chi,\chi(\mathbf{k}'),\varphi(\mathbf{l}'),\varphi(\mathbf{k}',\mathbf{l}')^*,$$

where the λ-measures are as above. The state vectors are indeed complete! Since the verification is similar for all the elements of M, we shall only treat the "1,1" element in detail. We want to verify that

$$1-Z=\chi_\Lambda\chi_\Lambda{}^*+\int d\mathbf{q}\,\chi_\xi\chi_\xi{}^*+\tfrac{1}{2}\iint d\mathbf{q}_1d\mathbf{q}_2\chi_{\xi_1,\xi_2}\chi_{\xi_1,\xi_2}{}^*$$
$$=\frac{2\gamma(\Lambda)}{\eta'(\Lambda)}+1-Z+2Z(I+J+K)$$

with

$$I=\iint d\mathbf{q}_1d\mathbf{q}_2\,|g(\mathbf{q}_1)|^2|g(\mathbf{q}_2)|^2/\eta^+(\xi_1+\xi_2)\eta^-(\xi_1+\xi_2),$$
$$J=2Z\int d\mathbf{q}\,|g(\mathbf{q})|^2/\eta^+(\xi+M)\eta^+(\xi+M),$$
$$K=-\int d\mathbf{q}\,|g(\mathbf{q})|^2\{[\eta^+(\xi+M)]^{-1}+[\eta^-(\xi+M)]^{-1}\}.$$

To do this, we need some properties of the $V\theta$ spectral function $\eta(\chi+\mu)=Z+\gamma(\chi+\mu)\alpha(\chi+\mu-M)$: Recall, first, that the $N\theta$ spectral function $\alpha(\chi)$ in the complex χ plane has a cut along $\mu\le\chi<\infty$, vanishes only at $\chi=M$ (it has no complex zeros), $\alpha(\chi)<0$ for real $\chi<M$, and $\alpha(\chi)$ increases monotonically as χ increases for $\chi<\mu$. Thus, from the spectral representation for $\gamma(\chi)$, we find $\gamma(\chi)$ is real, negative, and decreases monotonically as χ increases for $\chi<M+\mu$. Note also

$$\gamma(\chi)=Z\int d\mathbf{k}\,\frac{|g(\mathbf{k})|^2}{(\chi-M-\omega)}+\iint d\mathbf{k}d\mathbf{l}\,\frac{|g(\mathbf{k})|^2|g(\mathbf{l})|^2}{(\chi-\omega-\nu)},$$

so that it has branch points at $\chi=M+\mu, 2\mu$. Next we consider the real zeros of $\eta(\chi+\mu)$: (i) For $2M-\mu<\chi<M$, $\gamma(\chi+\mu)$ is negative and decreases monotonically and $\alpha(\chi+\mu-M)$ starts at zero, becomes positive, and increases monotonically; hence, there can be at most one zero of $\eta(\chi+\mu)$ in this interval and it occurs iff $|\gamma(M+\mu)|>Z/\alpha(\mu)$. That is, the $V_\gamma\theta$ bound state does not exist if this is not satisfied; (ii) for $\chi < 2M-\mu$, both $\gamma(\chi+\mu)$ and $\alpha(\chi+\mu-M)$ are negative so that $\eta(\chi+\mu) \geq Z$. Also $\eta(\chi)$ has no complex zeros:

$$0 = \eta(\chi) - \eta(\chi)^* = \int d\mathbf{k}\,|g(\mathbf{k})|^2 \times \left(\frac{\alpha(\chi - M)}{\alpha(\chi-\omega)} - \frac{\alpha(\chi^* - M)}{\alpha(\chi^* - \omega)}\right)$$

so $\forall\omega$ such that $\mu < \omega < \infty$

$$\frac{\alpha(\chi - M)}{\alpha(\chi-\omega)} - \frac{\alpha(\chi^* - M)}{\alpha(\chi^* - \omega)} = 0;$$

and, taking $|\omega| \to \infty$, $\mathrm{Im}\,\alpha(\chi - M) = 0$ so that χ is real. Finally, to display the cuts, we write

$$\eta(\chi + \mu) = \alpha(\chi + \mu - M)\rho(\chi+\mu) + 2Z,$$

where

$$\rho(\chi+\mu) \equiv \gamma(\chi+\mu) - \frac{Z}{\alpha(\chi+\mu-M)} = -\frac{Z^2}{\chi+\mu-2M} + \iint d\mathbf{k}\,d\mathbf{l}\,\frac{|g(\mathbf{k})|^2|g(\mathbf{l})|^2}{\chi+\mu-\omega-\nu}$$

so that $\alpha(\chi+\mu-M)$ has the $V_\gamma\theta$ cut along $M < \chi < \infty$ and $\rho(\chi+\mu)$ the $N\theta\theta$ cut along $\mu < \chi < \infty$.

Hence,

$$I = \frac{1}{2\pi i}\int_\mu^\infty d\chi\,\frac{1}{\alpha^+(\chi+\mu-M)}\left(\frac{1}{\eta^+(\chi+\mu)} - \frac{1}{\eta^-(\chi+\mu)}\right)$$

$$J + K = \frac{1}{2\pi i}\int_M^\infty d\chi\,\frac{1}{2}\left(\frac{1}{\alpha^+(\chi+\mu-M)} + \frac{1}{\alpha^-(\chi+\mu-M)}\right)\times\left(\frac{1}{\eta^+(\chi+\mu)} - \frac{1}{\eta^-(\chi+\mu)}\right)$$
$$+ \frac{1}{2\pi i}\int_M^\infty dk\,\frac{1}{2}\left(\frac{1}{\eta^+(\chi+\mu)} + \frac{1}{\eta^-(\chi+\mu)}\right)\times\left(\frac{1}{\alpha^+(\chi+\mu-M)} - \frac{1}{\alpha^-(\chi+\mu-M)}\right).$$

These cancel the cuts of the contour integral

$$\frac{1}{2\pi i}\int_C d\chi\,\frac{1}{\alpha(\chi+\mu-M)}\,\frac{1}{\eta(\chi+\mu)} = 0,$$

where C is the clockwise, simple closed contour about the entire χ complex plane (C excludes all the singularities). The pole at $\chi = -\mu$ cancels the contribution as $|\chi| \to \infty$, and the pole at $\chi = \Lambda - \mu$ cancels the $2\gamma(N)/\eta'(N)$ term in the "1, 1" element equation; thus it is verified.

ACKNOWLEDGMENTS

The author wishes to thank Professor E.C.G. Sudarshan for several valuable discussions and for his criticism of the manuscript. He also wishes to acknowledge the stimulating atmosphere and warm hospitality of the Center for Particle Theory of the University of Texas at Austin in the summer of 1971 where this work was begun.

1 M. Bolsterli, Phys. Rev. **166**, 1760 (1968).
2 G. Källén and W. Pauli, Kgl. Danske Videnshab. Selskab, Mat. Fys. Medd. **30**, No. 7 (1955); W. Heisenberg, Nucl. Phys. **4**, 532 (1957).
3 Boldface Latin letters label 3-momentum and Greek letters the corresponding energy. Used here are $\omega(\mathbf{k}) = (\mu^2 + \mathbf{k}^2)$, $\nu(\mathbf{l})$, $\sigma(\mathbf{s})$, $\tau(\mathbf{t})$, $\chi(\mathbf{p})$, $\xi(\mathbf{q})$, $\mu(\mathbf{m})$, $\delta(\mathbf{g})$.
4 A. Pagnamenta, J. Math Phys. **6**, 955 (1965); **7**, 356 (1966).

Incoherent Exciton Quenching on Lattices*

Richard P. Hemenger and Robert M. Pearlstein
Biology Division, Oak Ridge National Laboratory,† Oak Ridge, Tennessee 37830

and

Katja Lakatos-Lindenberg‡
Department of Chemistry, University of California at San Diego, La Jolla, California 92037

(Received 27 October 1971; Revised Manuscript Received 3 March 1972)

The time-dependent excitation function $\Phi(t)$ for exciton quenching in lattices of any dimension is derived from a general master equation in terms of perfect lattice Green's functions. Attention is largely restricted to nearest-neighbor interactions for which the Green's functions have simple forms. The quenchers are characterized by three dimensionless rate parameters: λ, for nearest neighbor host-quencher energy transfer; μ, for back transfer from quencher to host; and Q, for irreversible degradation of excitation on the quencher. Simple expressions for the Laplace transform of $\Phi(t)$ under two different initial conditions are given for low concentrations of periodically placed quenchers that have identical but arbitrary λ, μ, and Q. Randomly placed quenchers are treated by adapting the coherent potential approximation; in three dimensions, this method gives a $\Phi(t)$ that is identical to that with periodic quenchers of the same low concentration. Lattices with two types of defects—one with $\mu = 0$ and one with $\mu > 0$, a case of particular interest in organic crystals—are treated in some detail. Finally, it is shown that energy transfer anisotropies have little effect on $\Phi(t)$, unless the smallest relative transfer rate is at least as small as the quencher concentration.

1. INTRODUCTION

The quenching of migratory excited states of molecular aggregates is an important process in organic crystals,[1] photosynthetic units,[2] and a variety of polymers.[3] For incoherent excitons, this process is described by a master equation[4] which, for practical purposes, is equivalent to a random-walk formulation.[5,6] In this paper, we extend our earlier one-dimensional results[5] to two- and three-dimensional systems and generalize our previous description of the quenching interaction. Previously,[5] we distinguished two types of quencher, i.e., a defect which irreversibly removes excitation from the lattice. We the $V_r\theta$ bound state does not exist if this is not satisfied; (ii) for $\chi < 2M-\mu$, both $\gamma(\chi+\mu)$ and $\alpha(\chi+\mu-M)$ are negative so that $\eta(\chi+\mu) \geq Z$. Also $\eta(\chi)$ has no complex zeros:

$$0 = \eta(\chi) - \eta(\chi)^* = \int d\mathbf{k}\,|g(\mathbf{k})|^2 \times \left(\frac{\alpha(\chi - M)}{\alpha(\chi-\omega)} - \frac{\alpha(\chi^* - M)}{\alpha(\chi^*-\omega)}\right)$$

so $\forall\omega$ such that $\mu < \omega < \infty$

$$\frac{\alpha(\chi - M)}{\alpha(\chi-\omega)} - \frac{\alpha(\chi^* - M)}{\alpha(\chi^*-\omega)} = 0;$$

and, taking $|\omega| \to \infty$, $\mathrm{Im}\,\alpha(\chi - M) = 0$ so that χ is real. Finally, to display the cuts, we write

$$\eta(\chi+\mu) = \alpha(\chi+\mu-M)\rho(\chi+\mu) + 2Z,$$

where

$$\rho(\chi+\mu) \equiv \gamma(\chi+\mu) - \frac{Z}{\alpha(\chi+\mu-M)} = -\frac{Z^2}{\chi+\mu-2M} + \iint d\mathbf{k}d\mathbf{l}\,\frac{|g(\mathbf{k})|^2|g(\mathbf{l})|^2}{\chi+\mu-\omega-\nu}$$

so that $\alpha(\chi+\mu-M)$ has the $V_r\theta$ cut along $M < \chi < \infty$ and $\rho(\chi+\mu)$ the $N\theta\theta$ cut along $\mu < \chi < \infty$.

Hence,

$$I = \frac{1}{2\pi i}\int_\mu^\infty d\chi\,\frac{1}{\alpha^+(\chi+\mu-M)}\left(\frac{1}{\eta^+(\chi+\mu)} - \frac{1}{\eta^-(\chi+\mu)}\right)$$

$$J + K = \frac{1}{2\pi i}\int_M^\infty d\chi\,\frac{1}{2}\left(\frac{1}{\alpha^+(\chi+\mu-M)} + \frac{1}{\alpha^-(\chi+\mu-M)}\right) \times\left(\frac{1}{\eta^+(\chi+\mu)} - \frac{1}{\eta^-(\chi+\mu)}\right) + \frac{1}{2\pi i}\int_M^\infty dk\,\frac{1}{2}\left(\frac{1}{\eta^+(\chi+\mu)} + \frac{1}{\eta^-(\chi+\mu)}\right) \times\left(\frac{1}{\alpha^+(\chi+\mu-M)} - \frac{1}{\alpha^-(\chi+\mu-M)}\right).$$

These cancel the cuts of the contour integral

$$\frac{1}{2\pi i}\int_C d\chi\,\frac{1}{\alpha(\chi+\mu-M)}\,\frac{1}{\eta(\chi+\mu)} = 0,$$

where C is the clockwise, simple closed contour about the entire χ complex plane (C excludes all the singularities). The pole at $\chi = -\mu$ cancels the contribution as $|\chi| \to \infty$, and the pole at $\chi = \Lambda - \mu$ cancels the $2\gamma(N)/\eta'(N)$ term in the "1, 1" element equation; thus it is verified.

ACKNOWLEDGMENTS

The author wishes to thank Professor E.C.G. Sudarshan for several valuable discussions and for his criticism of the manuscript. He also wishes to acknowledge the stimulating atmosphere and warm hospitality of the Center for Particle Theory of the University of Texas at Austin in the summer of 1971 where this work was begun.

1 M. Bolsterli, Phys. Rev. 166, 1760 (1968).
2 G. Källén and W. Pauli, Kgl. Danske Videnshab. Selskab, Mat. Fys. Medd. 30, No. 7 (1955); W. Heisenberg, Nucl. Phys. 4, 532 (1957).
3 Boldface Latin letters label 3-momentum and Greek letters the corresponding energy. Used here are $\omega(\mathbf{k}) = (\mu^2 + \mathbf{k}^2)$, $\nu(\mathbf{l})$, $\sigma(\mathbf{s})$, $\tau(\mathbf{t})$, $\chi(\mathbf{p})$, $\xi(\mathbf{q})$, $\mu(\mathbf{m})$, $\delta(\mathbf{g})$.
4 A. Pagnamenta, J. Math Phys. 6, 955 (1965); 7, 356 (1966).

Incoherent Exciton Quenching on Lattices*

Richard P. Hemenger and Robert M. Pearlstein
Biology Division, Oak Ridge National Laboratory,† Oak Ridge, Tennessee 37830

and

Katja Lakatos-Lindenberg‡
Department of Chemistry, University of California at San Diego, La Jolla, California 92037

(Received 27 October 1971; Revised Manuscript Received 3 March 1972)

The time-dependent excitation function $\Phi(t)$ for exciton quenching in lattices of any dimension is derived from a general master equation in terms of perfect lattice Green's functions. Attention is largely restricted to nearest-neighbor interactions for which the Green's functions have simple forms. The quenchers are characterized by three dimensionless rate parameters: λ, for nearest neighbor host-quencher energy transfer; μ, for back transfer from quencher to host; and Q, for irreversible degradation of excitation on the quencher. Simple expressions for the Laplace transform of $\Phi(t)$ under two different initial conditions are given for low concentrations of periodically placed quenchers that have identical but arbitrary λ, μ, and Q. Randomly placed quenchers are treated by adapting the coherent potential approximation; in three dimensions, this method gives a $\Phi(t)$ that is identical to that with periodic quenchers of the same low concentration. Lattices with two types of defects—one with $\mu = 0$ and one with $\mu > 0$, a case of particular interest in organic crystals—are treated in some detail. Finally, it is shown that energy transfer anisotropies have little effect on $\Phi(t)$, unless the smallest relative transfer rate is at least as small as the quencher concentration.

1. INTRODUCTION

The quenching of migratory excited states of molecular aggregates is an important process in organic crystals,[1] photosynthetic units,[2] and a variety of polymers.[3] For incoherent excitons, this process is described by a master equation[4] which, for practical purposes, is equivalent to a random-walk formulation.[5,6] In this paper, we extend our earlier one-dimensional results[5] to two- and three-dimensional systems and generalize our previous description of the quenching interaction. Previously,[5] we distinguished two types of quencher, i.e., a defect which irreversibly removes excitation from the lattice. We defined a nondisruptive quencher to be one whose presence does not alter the interactions of the quenched host sites with their neighbors [Fig. 1(a)], for example, metal–ion quenching of triplet excitons in polynucleotides.[5] A disruptive quencher was defined to be essentially a substitutional impurity [Fig. 1(b)].[5] Here, we generalize our concept of quenching as shown in Fig. 1(c). The symbol $\otimes$ in Fig. 1(c) does not necessarily represent an interstitial site, but rather a quenching rate process, for example, the differential fluorescence rate of a substitutional impurity in an organic crystal.[7] The generalized quencher is nondisruptive with respect to the irreversible removal of excitation from the lattice, but is disruptive in that it changes the transfer rate constants with the nearest neighbors of the quenched site. In this sense, a generalized quencher is analogous, in the theory of lattice dynamics, to a substitutional impurity which has both a different mass and different nearest-neighbor force constants from those of the host sites. The most significant feature of this generalized quencher is the inclusion of an arbitrary back transfer rate μ. Recent evidence suggests that some form of excitation feedback from host traps strongly influences the exciton quenching kinetics in doped polyacenes[7] and in polynucleotides.[8]

In Sec. 2, we express the total excitation at any time on a periodically quenched lattice in terms of Green's functions for a lattice without quenchers. Except for purely nondisruptive quenchers (Fig. 1(a) or Fig. 1(c) with $\lambda = \mu = 1$), only nearest-neighbor interactions are included. We also obtain a generalization of the coherent potential approximation[5,9] with which we treat generalized quenchers that have random locations and random back-transfer rates.

In Sec. 3A we give results for periodically quenched lattices in one, two, and three dimensions for two different initial conditions. The case of partially reversible quenching is considered in some detail. In Sec. 3B results for random quencher locations are given for two- and three-dimensional lattices. We also consider the detailed time dependences of host and irreversible impurities ($\mu = 0$) in the presence of a second type of defect with quenching parameter $\mu > 0$. In Sec. 3C the effects of certain energy transfer anisotropies are calculated.

In the final section we summarize the results and remark briefly on exciton quenching with extended interactions.

2. SOLUTIONS FOR SHORT-RANGE QUENCHING IN TERMS OF GREEN'S FUNCTIONS

We consider a variety of exciton quenching problems all of which are described by the general master equation

$$F^{-1}\dot{\rho}_1(t) = \sum_{1' \neq 1} \Omega_{11'}[\rho_{1'}(t) - \rho_1(t)] + \sum_{1'} \Delta_{11'}\rho_{1'}(t), \tag{2.1}$$

where $\rho_1(t)$ is the probability that site $\mathbf{1}$ is occupied at time t. F is the reversible pairwise rate constant for exciton transfer between nearest-neighboring nonquenchers. $F\Omega_{11'}$ is defined as the reversible pairwise rate constant between sites $\mathbf{1}$ and $\mathbf{1}'$ in the absence of quenchers. $F\Delta_{11'}$ is defined such that, for $\mathbf{1} \neq \mathbf{1}'$, $F(\Omega_{11'} + \Delta_{11'})$ is the (not necessarily reversible) pairwise rate constant from site $\mathbf{1}'$ to site $\mathbf{1}$ in the presence of quenchers. $F\Delta_{11}$ is defined such that $F(\sum_{1' \neq 1}\Omega_{11'} - \Delta_{11})$ is the rate at which excitation leaves site $\mathbf{1}$ in the presence of quenchers. Two considerations dictate the choice of notation in Eq. (2.1). Separation of terms involving the Ω's and the Δ's is convenient for our Green's function formalism. The term in the Δ's is the simplest way to include all features of generalized quenching.

We use the dimensionless Laplace transform

$$\tilde{\rho}_1(s) \equiv F \int_0^\infty e^{-sFt} \rho_1(t)dt. \tag{2.2}$$

By transforming Eq. (2.1), one obtains

$$-(s + q_1)\tilde{\rho}_1(s) + \sum_{1' \neq 1} \Omega_{11'}\tilde{\rho}_{1'}(s) + \sum_{1'} \Delta_{11'}\tilde{\rho}_{1'}(s) = -\rho_1^0, \tag{2.3}$$

where $\rho_1^0 \equiv \rho_1(t = 0)$ and $q_1 \equiv \sum_{1' \neq 1} \Omega_{11'}$.

Defining the Green's function of the perfect lattice by

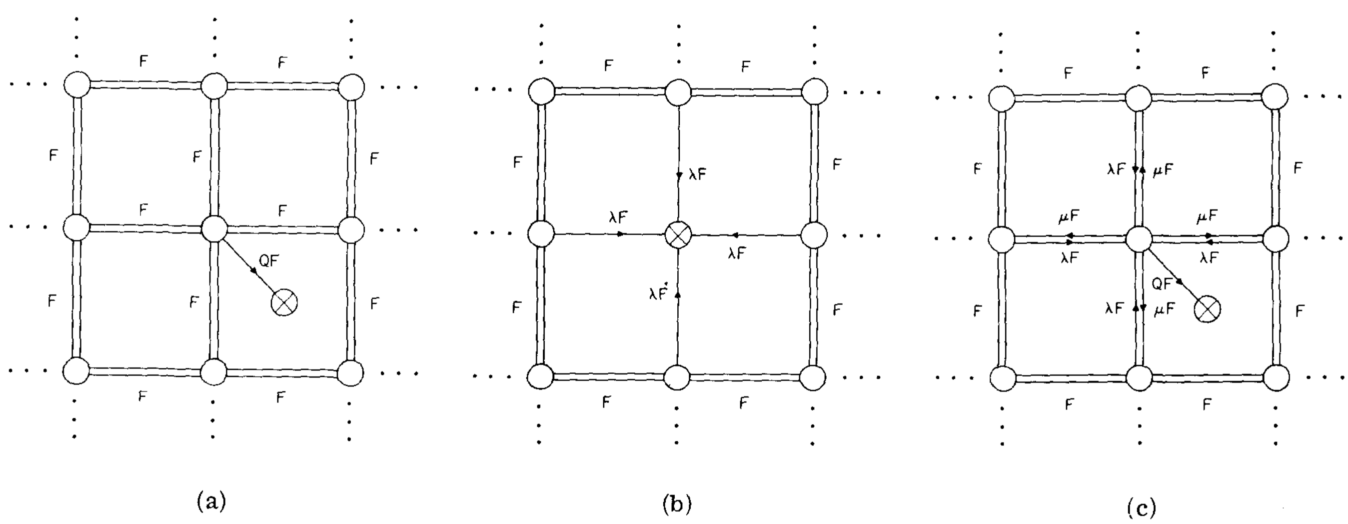

FIG. 1. Various types of quenching on a two-dimensional square lattice. ○, Host site; ⊗, quenching site (not necessarily an interstitial impurity—see text); ═, symmetric exciton pathway; —▶—, asymmetric exciton pathway. For simplicity, only nearest neighbor pathways are indicated: (a) purely nondisruptive quencher; (b) purely disruptive quencher; (c) generalized quencher.

$$-(s+q_{\mathbf{l}})g_{\mathbf{l}\mathbf{l}'}(s) + \sum_{\mathbf{l}''\neq \mathbf{l}} \Omega_{\mathbf{l}\mathbf{l}''}g_{\mathbf{l}''\mathbf{l}'}(s) = \delta_{\mathbf{l}\mathbf{l}'}, \tag{2.4}$$

where $\delta_{\mathbf{l}\mathbf{l}'}$ is the Kronecker delta, one obtains

$$\tilde{\rho}_{\mathbf{l}}(s) = -\sum_{\mathbf{l}'\mathbf{l}''} g_{\mathbf{l}\mathbf{l}''}(s)\,[\rho^0_{\mathbf{l}''}\delta_{\mathbf{l}''\mathbf{l}'} + \Delta_{\mathbf{l}''\mathbf{l}'}\tilde{\rho}_{\mathbf{l}'}(s)]. \tag{2.5}$$

The quantities of greatest interest are the lattice excitation function

$$\Phi(t) \equiv \sum_{\mathbf{l}} \rho_{\mathbf{l}}(t), \tag{2.6}$$

and its moments, which are given by

$$M_m(\Phi) \equiv \int_0^\infty t^m\Phi(t)\,dt = \frac{(-1)^m}{F^{m+1}}\left[\frac{d^m}{ds^m}\tilde{\Phi}(s)\right]_{s=0}, \tag{2.7}$$

where

$$\tilde{\Phi}(s) = \sum_{\mathbf{l}}\tilde{\rho}_{\mathbf{l}}(s). \tag{2.8}$$

The simplest quenching problems involve primitive n-dimensional lattices in which the quenching interaction is only of nearest-neighbor type. In Sec. 2A we treat primitive lattices with periodic short-range quenchers. For purely nondisruptive quenching we consider arbitrary host–host interactions. For generalized quenching we restrict ourselves to nearest-neighbor interactions. At the end of Sec. 2A a finite host lifetime and quenching due to the finite impurity lifetime are introduced. In Sec. 2B we develop a formalism for treating random quenching.

A. Periodic Quenching

We consider a lattice of N sites with periodic boundary conditions and a single generalized quencher at the origin [Fig. 1(c)]. As far as $\Phi(t)$ is concerned, such a finite lattice with any initial condition is equivalent to an infinite lattice (having periodic quenchers of quencher concentration $c = 1/N$) with an appropriately chosen initial condition.[10]

For this case we have [see Fig. 1(c) for notation]

$$\Delta_{\mathbf{l}\mathbf{l}'} = \{(\lambda-1)(\delta_{\mathbf{l}0}-\delta_{\mathbf{l}\mathbf{m}})\delta_{\mathbf{l}'\mathbf{m}} + (\mu-1)\delta_{\mathbf{l}\mathbf{m}}\delta_{\mathbf{l}'0} - [Q+q(\mu-1)]\delta_{\mathbf{l}0}\delta_{\mathbf{l}'0}\}, \tag{2.9}$$

where $\mathbf{m}$ is any of the nearest neighbors of $\mathbf{l}=0$, and q is the coordination number of the lattice. A rate constant τ^{-1} for a uniform (independent of $\mathbf{l}$) de-excitation process such as fluorescence can be included in $\Delta_{\mathbf{l}\mathbf{l}}$. However, since the effect of any such process can be included simply by multiplying the result for $\Phi(t)$ by $\exp(-t/\tau)$ [see, e.g., Eq. (2.21)], we have omitted such a term from Eq. (2.9). To illustrate how the terms in Eq. (2.9) are obtained, we explicitly construct Δ_{00}. Because we have assumed that the quencher affects only the interactions with its own nearest neighbors and leaves more distant interactions unchanged, it follows from the definition of $\Delta_{\mathbf{l}\mathbf{l}}$ that

$$\sum_{|\mathbf{l}'|=1}\Omega_{0\mathbf{l}'} - \Delta_{00} = Q + q\mu,$$

and since $\sum_{|\mathbf{l}'|=1}\Omega_{0\mathbf{l}'} = q$,

$$\Delta_{00} = -Q - q(\mu-1).$$

Other contributions to $\Delta_{\mathbf{l}\mathbf{l}'}$ are evaluated similarly. From Eqs. (2.5) and (2.8),[11]

$$\tilde{\Phi}(s) = s^{-1}[1 - Q\tilde{\rho}_0(s)], \tag{2.10}$$

where

$$\tilde{\rho}_0 = \frac{qC(\lambda-1)(g_0-g_1) - H_0[1+(\lambda-1)(qg_1-B)]}{\{1-Qg_0+(\lambda-1)[qg_1-B-Q(qg_1^2-g_0B)]-(\mu-1)q(g_0-g_1)\}} \tag{2.11}$$

and where we have used the relation $\sum_{\mathbf{l}} g_{\mathbf{l}\mathbf{l}'} = -s^{-1}$. (This result is obtained by taking the Laplace transform of the conservation-of-probability equation $\sum_{\mathbf{l}}\rho_{\mathbf{l}}(t)=1$, which holds for a perfect lattice regardless of initial condition.) In Eq. (2.11), $g_n \equiv g_{\mathbf{l}\mathbf{l}'}$ for $|\mathbf{l}-\mathbf{l}'| = n$,

$$B(s) \equiv \frac{1}{q}\sum_{|\mathbf{l}|=1}\sum_{|\mathbf{l}'|=1} g_{\mathbf{l}\mathbf{l}'}(s), \tag{2.12}$$

and

$$C(s) \equiv q\sum_{|\mathbf{l}|=1} H_{\mathbf{l}}(s) \tag{2.13}$$

with

$$H_{\mathbf{l}}(s) \equiv \sum_{\mathbf{l}'} g_{\mathbf{l}\mathbf{l}'}(s)\rho^0_{\mathbf{l}'}. \tag{2.14}$$

The purely nondisruptive quencher ($\mu=\lambda=1$) can be treated very generally. From Eqs. (2.10) and (2.11),

$$\tilde{\Phi}(s) = \frac{1}{s}\left(1 + \frac{QH_0(s)}{1-Qg_0(s)}\right). \tag{2.15}$$

From Eqs. (2.7) and (2.15), the zeroth moment of $\Phi(t)$ is

$$FM_0 = \tilde{\Phi}(0) = \frac{N}{Q} - N\left(f_{00} - \sum_{\mathbf{l}} f_{0\mathbf{l}}\rho^0_{\mathbf{l}}\right), \tag{2.16}$$

where[12]

$$f_{\mathbf{l}\mathbf{l}'} \equiv \lim_{s\to 0}[g_{\mathbf{l}\mathbf{l}'}(s) + (1/Ns)]. \tag{2.17}$$

Equation (2.16) shows that for primitive lattices of any dimension with arbitrary host–host interactions and any initial condition, the short-range, purely nondisruptive quenching of incoherent excitons can be described as an additive sequence of events, with a mean time to reach the nearest neighbor of the quencher plus one to be quenched from that neighbor, as already noted for certain one-dimensional situations.[10]

If Q, λ, and μ are arbitrary, Eq. (2.11) greatly simplifies for host–host interactions restricted to those between nearest neighbors, i.e., $\Omega_{\mathbf{l}\mathbf{l}'} = \delta_{|\mathbf{l}-\mathbf{l}'|,1}$. Here,

$$\tilde{\rho}_0 = \frac{(\lambda-1)(1+sg_0)\rho^0_0 + \lambda qH_0}{(\lambda-1)(s+Q)(1+sg_0) - \mu q(1+sg_0) + \lambda q(s+Q)g_0}, \tag{2.18}$$

where we have used Eq. (2.3). Using Eqs. (2.7), (2.10), and (2.18), one obtains

$$FM_0 = \left[\frac{1}{Q}\left(1 + \frac{\mu}{\lambda}(N-1)\right) + \frac{(N-1)}{\lambda q}(1-\rho_0^0)\right] - \left[N\left(f_{00} - \sum_{\mathbf{l}} f_{0\mathbf{l}}\rho_{\mathbf{l}}^0\right) + \frac{(N-1)}{q}(1-\rho_0^0)\right]. \quad (2.19)$$

Equation (2.19) shows for a generalized quencher with only nearest-neighbor host–host interactions what Eq. (2.16) does for purely nondisruptive quenching and arbitrary interactions. Equations (2.18) and (2.19) describe exciton quenching with an arbitrary amount of back transfer. They include the case of purely disruptive quenching, $Q = \infty$.

In an important class of exciton quenching phenomena[13] the parameter of irreversibility Q arises solely from the finite lifetime of the quenching impurities. If one neglects other sources of quenching, such as that due to host traps,[1] then in Eq. (2.18)

$$FQ = \tau_Q^{-1} - \tau_H^{-1}, \quad (2.20)$$

where τ_Q and τ_H are the finite lifetimes of the quencher and host, respectively. The right-hand side of Eq. (2.20) contains a difference of rates because the fluorescence rate of the host has been transformed out uniformly for all sites. For impulsive excitation at $t = 0$, the time dependence of the host and quencher excitations are then given, respectively, by

$$P_H(t) = e^{-t/\tau_H}[\Phi(t) - \rho_0(t)], \quad (2.21a)$$

$$P_Q(t) = e^{-t/\tau_H}\rho_0(t). \quad (2.21b)$$

B. Random Quenching

We distinguish two types of randomness in the quenching process. In one type the quenchers are randomly located; in the other, the values of one or more of the quenching parameters (λ, μ, Q) are randomly distributed. Both types can be treated by extending the coherent potential approximation[9] (CPA). A CPA for randomly located nondisruptive quenchers in one dimension has already been given.[5]

In the usual CPA a self-consistently calculated potential replaces the true potential or quencher at every site but one. This approach works well if the lattice disorder is purely diagonal, e.g., for purely nondisruptive quenching in any number of dimensions. However, randomly placed generalized quenchers introduce both diagonal and off-diagonal disorder. In this case a difficulty arises because generalized quenchers cannot be placed at adjacent sites. To avoid this difficulty we introduce a model with two sublattices A and B. In this model we artificially distinguish each site from its nearest neighbors and assume that generalized quenchers are randomly distributed only over sublattice A. Such a "diatomic" lattice approximates the true situation extremely well at low quencher concentration. (This model is analogous to one that would be appropriate for lattice vibrations on a lattice with a low concentration of mass and force constant defects.) We calculate a Green's function $\bar{g}_{\mathbf{l}\mathbf{l}'}$ for the situation in which there is a generalized quencher with parameters $\bar{\lambda}, \bar{\mu}$, and $\bar{Q}$ at every site of sublattice A. These parameters are then determined self-consistently as usual in the CPA.

The Green's function satisfies

$$-(s + q\bar{\mu} + \bar{Q})\bar{g}_{\mathbf{l}\mathbf{l}'} + \bar{\lambda} \sum_{|\mathbf{l}''-\mathbf{l}|=1} \bar{g}_{\mathbf{l}''\mathbf{l}'} = \delta_{\mathbf{l}\mathbf{l}'}, \quad \mathbf{l} \in \mathrm{A}, \quad (2.22a)$$

$$-(s + \bar{\lambda}q)\bar{g}_{\mathbf{l}\mathbf{l}'} + \bar{\mu} \sum_{|\mathbf{l}''-\mathbf{l}|=1} \bar{g}_{\mathbf{l}''\mathbf{l}'} = \delta_{\mathbf{l}\mathbf{l}'}, \quad \mathbf{l} \in B. \quad (2.22b)$$

The solution of Eq. (2.22) for simple cubic lattices of any dimension is

$$\bar{g}_{\mathbf{l}\mathbf{l}'}(s) = \frac{1}{N}\sum_{\mathbf{k}} \frac{e^{2\pi i\mathbf{k}\cdot(\mathbf{l}-\mathbf{l}')/N^{1/3}} a_{\mathbf{l}\mathbf{l}'}(\mathbf{k})}{\bar{\lambda}\bar{\mu}\theta_{\mathbf{k}}^2 - (s + \bar{\lambda}q)(s + \bar{\mu}q + \bar{Q})}, \quad (2.23)$$

where $\mathbf{k} = (k_1, k_2, \ldots, k_{q/2})$,

$$\theta_{\mathbf{k}} = 2\sum_{n=1}^{q/2} \cos\left(\frac{2\pi k_n}{N^{1/3}}\right) \quad (2.24)$$

and

$$a_{\mathbf{l}\mathbf{l}'}(\mathbf{k}) = \begin{cases} s + \bar{\lambda}q, & \mathbf{l} \in A, \quad \mathbf{l}' \in A, \\ \bar{\lambda}\theta_{\mathbf{k}}, & \mathbf{l} \in A, \quad \mathbf{l}' \in B, \\ \bar{\mu}\theta_{\mathbf{k}}, & \mathbf{l} \in B, \quad \mathbf{l}' \in A, \\ s + \bar{\mu}q + \bar{Q}, & \mathbf{l} \in B, \quad \mathbf{l}' \in B. \end{cases} \quad (2.25)$$

For an arbitrary distribution of generalized quenchers with parameters $\lambda_{\mathbf{l}}, \mu_{\mathbf{l}}$, and $Q_{\mathbf{l}}$ on the A sublattice, each of which replaces a coherent quencher (see Fig. 2),

$$\tilde{\rho}_{\mathbf{l}}(s) = -\bar{H}_{\mathbf{l}}(s) - \sum_{\mathbf{l}'\mathbf{l}''} \bar{g}_{\mathbf{l}\mathbf{l}''}(s)\bar{\Delta}_{\mathbf{l}''\mathbf{l}'}\tilde{\rho}_{\mathbf{l}'}(s), \quad (2.26)$$

where for a uniform initial condition

$$\bar{H}_{\mathbf{l}}(s) = \sum_{\mathbf{l}'} \bar{g}_{\mathbf{l}\mathbf{l}'}(s)\rho_{\mathbf{l}'}^0 = \frac{-1}{2N}\frac{[2s + 2q(1+\bar{\mu}) + \bar{Q}]}{\{s^2 + s[q(1+\bar{\mu}) + \bar{Q}] + q\bar{Q}\}} \quad (2.27)$$

and

$$\bar{\Delta}_{\mathbf{l}\mathbf{l}'} = \begin{cases} (\mu_{\mathbf{l}'} - \bar{\mu})\delta_{|\mathbf{l}-\mathbf{l}'|,1} - [Q_{\mathbf{l}'} - \bar{Q} + q(\mu_{\mathbf{l}'} - \bar{\mu})]\delta_{\mathbf{l}\mathbf{l}'}, & \mathbf{l}' \in A. \\ (\lambda_{\mathbf{l}'} - \bar{\lambda})[\delta_{|\mathbf{l}-\mathbf{l}'|,1} - \delta_{\mathbf{l}\mathbf{l}'}], & \mathbf{l}' \in B, \end{cases} \quad (2.28)$$

The function $\tilde{\Phi}$ is insensitive to the values of any of the λ's and μ's except for those which are of order of the quencher concentration. Since we are not interested here in the case of small λ, there is thus no great loss in generality to take $\lambda_{\mathbf{l}'} = \bar{\lambda} = 1$, which we do below. However, because the case of small μ is of interest (see Sec. 3), we retain $\mu_{\mathbf{l}'}$ and hence $\bar{\mu}$ as variables.

One can define a T matrix,[9] $\bar{t}_{\mathbf{l}\mathbf{l}'}$, in terms of which Eq. (2.26) becomes

$$\tilde{\rho}_{\mathbf{l}} = -\bar{H}_{\mathbf{l}} - \sum_{\mathbf{i}\mathbf{j}} \bar{g}_{\mathbf{l}\mathbf{i}}\bar{t}_{\mathbf{i}\mathbf{j}}\bar{H}_{\mathbf{j}} - \sum_{\substack{\mathbf{i}\mathbf{j}\mathbf{m}\mathbf{n} \\ \mathbf{n}\neq\mathbf{j}}} \bar{g}_{\mathbf{l}\mathbf{i}}\bar{t}_{\mathbf{i}\mathbf{j}}\bar{g}_{\mathbf{j}\mathbf{m}}\bar{t}_{\mathbf{m}\mathbf{n}}\bar{H}_{\mathbf{n}} - \cdots. \quad (2.29)$$

Direct substitution shows that Eq. (2.29) is satisfied if

$$\bar{t}_{\mathbf{l}\mathbf{l}'}(s;\mu_{\mathbf{l}'} - \bar{\mu}, Q_{\mathbf{l}'} - \bar{Q}) = \begin{cases} \dfrac{-\bar{\Delta}_{\mathbf{l}\mathbf{l}'}}{1 + \sum_{\mathbf{l}''} \bar{g}_{\mathbf{l}'\mathbf{l}''}\bar{\Delta}_{\mathbf{l}''\mathbf{l}'}}, & \mathbf{l}' \in A, \\ 0, & \mathbf{l}' \in B. \end{cases} \quad (2.30)$$

The sum in the denominator of Eq. (2.30) can be expressed in terms of $\bar{g}_0^A \equiv \bar{g}_{11}$ for $1 \in A$ with the aid of Eqs. (2.22), (2.23), and (2.25). Thus for $1' \in A$,

$$\bar{t}_{11'} = \begin{cases} \dfrac{q(\mu_{1'} - \bar{\mu}) + Q_{1'} - \bar{Q}}{(\mu_{1'}/\bar{\mu})[1 + (s + \bar{Q})\bar{g}_0^A] - (s + Q_{1'})\bar{g}_0^A}, & 1 = 1', \\ \dfrac{-(\mu_{1'} - \bar{\mu})}{(\mu_{1'}/\bar{\mu})[1 + (s + \bar{Q})\bar{g}_0^A] - (s + Q_{1'})\bar{g}_0^A}, & |1 - 1'| = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (2.31)$$

If $\bar{\mu}$ and $\bar{Q}$ are determined from the conditions

$$\langle \bar{t}_{1+1',1} \rangle \equiv \frac{1}{N} \sum_{1} \bar{t}_{1+1',1} = 0 \quad (2.32)$$

for each $1'$, then

$$\langle \tilde{\rho}_1 \rangle \equiv \frac{1}{N} \sum_{1} \tilde{\rho}_1 \equiv \frac{1}{N} \tilde{\Phi} \simeq \frac{1}{2N} \frac{[2s + 2q(1 + \bar{\mu}) + \bar{Q}]}{\{s^2 + s[q(1 + \bar{\mu}) + \bar{Q}] + q\bar{Q}\}}. \quad (2.33)$$

[In the last expression in Eq. (2.33) we have neglected terms of $O(\bar{t}^4)$. Lower order terms in $\bar{t}$ vanish identically because of condition (2.32) and the lack of certain diagonal terms in Eq. (2.29).] This corresponds to placing a generalized quencher of parameters $\bar{\mu}$ and $\bar{Q}$ on every site of the A sublattice except the 1th, as shown schematically in Fig. 2.

For an infinite system Eq. (2.32) is identical to

$$\int d\mu' \int dQ' p(\mu', Q') \bar{t}_{10}(s; \mu' - \bar{\mu}, Q' - \bar{Q}) = 0 \quad (2.34)$$

for each 1, where site 0 is any site of sublattice A. In Eq. (2.34), $p(\mu', Q')$ is the distribution of probability that at an arbitrary site of sublattice A there is a quencher of parameters μ' and Q'. As an example we consider further in Sec. 3, the quenchers are randomly located, μ' is a random variable with distribution $h(\mu')$, and

$$p(\mu', Q') = 2ch(\mu')\delta(Q' - Q) + (1 - 2c)\delta(\mu' - 1)\delta(Q'), \quad (2.35)$$

where c is the concentration of generalized quenchers in the whole lattice. If the quenchers all have identical parameters,

$$h(\mu') = \delta(\mu' - \mu). \quad (2.36)$$

3. RESULTS FOR NEAREST NEIGHBOR INTERACTIONS

A. Periodically Quenched Lattices

We consider two different initial conditions that may

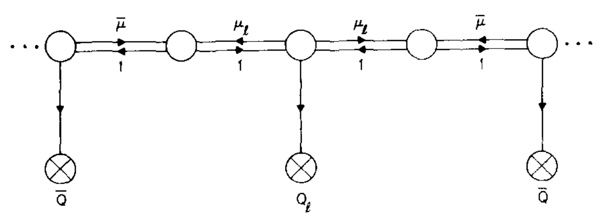

FIG. 2. Schematic diagram of generalized quencher arrangement in two-sublattice CPA.

be of interest with periodic quenching. Under uniform initial excitation, $\rho_1^0 = 1/N$ for all 1. If only the impurity sites are initially excited, $\rho_1^0 = \delta_{10}$. A third condition that may arise in practice is $\rho_1^0 = [1/(N-1)](1 - \delta_{10})$; i.e., all but the impurity sites are uniformly excited at $t = 0$. However, since the zero moments under the last condition are equivalent to those for the first condition to within $o(1/N)$, we do not treat the last condition separately, although it is easy to do so.

For any of these initial conditions, the only quantity required in Eq. (2.19) is f_{00}, which may be obtained from the relation

$$-qNf_{00} = qFM_0(Q = \infty, \lambda = 1, \mu) = \langle n \rangle, \quad (3.1)$$

where the first equality follows from Eq. (2.19) itself and M_0 is evaluated for the uniform initial condition. The quantity $\langle n \rangle$ is the mean first passage time for a random walk of constant stepping time $1/(qF)$.[6,14] Using values of $\langle n \rangle$ given by Montroll,[15,16] one obtains

$$-f_{00} = \begin{cases} (N^2 - 1)/12N, & \text{1D}, \\ (c_1/q)\log N + (c_2/q) + o(1/N), & \text{2D}, \\ (c_1/q) + o(1/N), & \text{3D}. \end{cases} \quad (3.2)$$

The values of the constants in Eq. (3.2) are collected in Table I. For the uniform initial condition, Eq. (2.19) becomes

$$FM_0 = \frac{1}{Q}\left(1 + \frac{\mu}{\lambda}(N - 1)\right) + \frac{(N-1)^2}{N\lambda q} - Nf_{00} - \frac{(N-1)^2}{Nq} \quad (3.3)$$

and, for the impurity initial condition,

$$FM_0 = (1/Q)[1 + (\mu/\lambda)(N - 1)]. \quad (3.4)$$

Note that FM_0 in Eq. (3.4) is independent of dimension and coordination number.

The lattice excitation function can be expressed as a sum of N exponentially decaying normal modes, each of which has a rate constant $\Gamma_k = -Fs_k$. The eigenvalues s_k are the poles of $\tilde{\Phi}(s)$ and are, therefore, the solutions of the secular equation

$$(\lambda - 1)(s + Q)\{[(N-1)/N] + sf(s)\}s - \mu q\{[(N-1)/N] + sf(s)\}s + \lambda q(s + Q)[(-1/N) + sf(s)] = 0, \quad (3.5)$$

where

$$f(s) \equiv g_0(s) + (Ns)^{-1}. \quad (3.6)$$

In general for large N, s_0 lies much closer to the origin than any of the other s_k. To evaluate s_0, in Eq. (3.5) it suffices to take $f(s) \simeq f_0$, a constant, and to retain only terms linear in s. The zero mode time constant is then

$$(\Gamma_0/F)^{-1} = -s_0^{-1} \simeq (1/Q)[1 + (\mu/\lambda)(N - 1)] + [(N-1)/\lambda q] - Nf_0 - [(N-1)/q]. \quad (3.7)$$

TABLE I. Values of constants in Eq. (3.2).

		q	c_1	c_2
2D (Ref. 16)	hexagonal	3	0.413 496 67	0.066 206 98
	square	4	0.318 309 89	0.195 056 17
	triangular	6	0.275 664 45	0.235 214 02
3D (Ref. 12)	SC	6	1.516 386 06	
	BCC	8	1.393 203 93	
	FCC	12	1.344 661 07	

In three dimensions $f_0 = f_{00} + o(\sqrt{1/N})$,[16] so that for a uniform initial condition [Eq. (3.3)],

$$-s_0^{-1} = FM_0[1 + o(\sqrt{1/N})]; \tag{3.8}$$

also

$$\operatorname{Res} \tilde{\Phi}(s)|_{s=s_0} = 1 - o(N^{-2/3}). \tag{3.9}$$

Equations (3.8) and (3.9) express the strong dominance of the zero mode in three dimensions with a uniform initial condition. Higher modes make a slightly greater contribution in two dimensions. In one dimension the higher modes continue to contribute even as $N \to \infty$.[10]

Back Transfer

Our description of the zero mode following Eq. (3.6) breaks down if $|Q| \sim |s_0|$, $\mu \sim |s_0|$, and $\lambda \gg |s_0|$, a set of conditions that may arise in organic crystals.[7] In this case it is necessary to retain terms of $o(s^2)$ in Eq. (35), with the result that there are two poles very close to the origin. In three dimensions,

$$\begin{aligned} 2N(\lambda - 1 + \lambda q f_{00}) s_{\substack{0\\1}} &= [\![-[(\lambda-1)QN + \lambda q[Qf_{00}N - 1) - \mu qN] \\ &\pm \{[(\lambda-1)QN + \lambda q(Qf_{00}N + 1)]^2 \\ &+ \mu^2 q^2 N^2 - 2\mu q(\lambda - 1)QN^2 \\ &- 2\lambda\mu q^2 N(Qf_{00}N - 1)\}^{1/2}]\!] \cdot [1 + o(\sqrt{1/N})]. \end{aligned} \tag{3.10}$$

Then from Eqs. (2.18), (2.21), and (3.10)

$$P_Q(t) = \lambda q e^{-t/\tau_H} \times \frac{[s_1 H_0(s_1) e^{Fs_1 t} - s_0 H_0(s_0) e^{Fs_0 t}]}{(\lambda - 1 + \lambda q f_{00})(s_1 - s_0)} [1 + o(N^{-2/3})]. \tag{3.11}$$

For the uniform initial condition,

$$sH_0(s) = -N^{-1} \tag{3.12}$$

and for the impurity initial condition,

$$sH_0(s) = -N^{-1} + sf_{00}. \tag{3.13}$$

Under either initial condition the effect of back transfer, if N is sufficiently large, is to decrease the rate of decay in the tail of $P_Q(t)$ and to increase $|dP_Q/dt|_{t=0}$ compared to their values when $\mu = 0$. Under the uniform initial condition, $P_Q(t)$ is a pulse whose maximum therefore occurs at an earlier time with back transfer than it does with $\mu = 0$.

B. Randomly Quenched Lattices

1. Random Locations

Combining Eqs. (2.31) and (2.34)–(2.36), one obtains two implicit equations for $\bar{\mu}$ and $\bar{Q}$ in terms of $\bar{g}_0(s; \bar{\mu}, \bar{Q})$. These are

$$\bar{\mu} = \frac{\mu - (s + Q)\bar{\mu}\bar{g}_0^A}{2c + (1 - 2c)\mu - (s + Q)\bar{\mu}\bar{g}_0^A}, \tag{3.14}$$

and

$$\bar{Q} = \frac{2cQ(1 - s\bar{\mu}\bar{g}_0^A)}{\mu - (s + Q)\bar{\mu}\bar{g}_0^A + 2c(1 - \mu + Q\bar{\mu}\bar{g}_0^A)}. \tag{3.15}$$

In three dimensions for $N \to \infty$,

$$\bar{\mu}\bar{g}_0^A(s; \bar{\mu}, \bar{Q}) = f_{00} + o\{[(s/2\bar{\mu})(1 + \bar{\mu} + \tfrac{1}{6}Q) + (\bar{Q}/2\bar{\mu})]^{1/2}\}. \tag{3.16}$$

Then from Eq. (2.33), for $s = o(c)$,

$$\tilde{\Phi}(s) = \left(\frac{(Q + s)f_{00} - \mu - c}{s^2 f_{00} + s(Qf_{00} - \mu - c) - cQ}\right)[1 + o(c^{2/3})]. \tag{3.17}$$

The result agrees to within $o(\sqrt{c}\,)$ with that of the corresponding periodic quencher case ($c = 1/N$), which is obtained from the substitution of Eq. (2.18) into Eq. (2.10) with the uniform initial condition and $\lambda = 1$. Thus, to lowest order in c, periodically and randomly quenched three-dimensional lattices are kinetically identical for all values of μ and Q. This conclusion is implicit in earlier work with $Q = \infty$ and $\mu = 1$.[17]

In two dimensions from Eqs. (2.7), (2.33), (3.14), and (3.15), to within $o(c)$,

$$FM_0 = (1/Q) + [\mu - Q\bar{\mu}\bar{g}_0^A(s = 0)]/cQ. \tag{3.18}$$

For $N \to \infty$,

$$\begin{aligned} \bar{\mu}\bar{g}_0^A(s = 0; \bar{\mu}, \bar{Q}) &= f(1/\bar{x})[1 + o(1/\bar{x})] \\ &= [-(1/4\pi)\ln(32\bar{x})][1 + o(1/\bar{x})], \end{aligned} \tag{3.19}$$

where

$$\bar{x} = (2\bar{\mu}/\bar{Q})_{s=0}. \tag{3.20}$$

Hence

$$FM_0 = (1/Q) + (1/4\pi c)[\ln D + \ln\ln D + \dots], \tag{3.21}$$

where

$$D = (8/\pi c)\exp(4\pi\mu/Q). \tag{3.22}$$

The leading term in Eq. (3.21), $Q^{-1} - (4\pi c)^{-1}\ln c$, is identical to that of the periodic case Eq. (3.3) with $c = 1/N$. Here, as in the periodic case, higher modes make a slightly greater contribution in two dimensions than in three.

In the one-dimensional case, treated in part elsewhere,[5] random and periodic cases have distinct kinetic properties.

2. Two Types of Defects

Up to this point we have dealt with only a single type of defect. In organic solids, however, there are frequently two types of defects—impurity quenchers and host traps—that influence excitation kinetics.[1]

The random location results show that for all physically realizable concentrations, there is no interference between defects, provided at least one type is irreversible. Hence one can treat each type of defect separately and superimpose the effects of both. Of particular interest is the case where the impurities have the fixed quenching parameters Q given by Eq. (2.20), $\lambda_Q \equiv \lambda$, and $\mu_Q = 0$ (irreversibility), and the traps have the parameters $Q_T, \lambda_T = 1$, and $\mu_T \equiv \mu$.

We proceed by solving the one-impurity and the one-trap problems separately. For the one-impurity problem with $\mu_Q = 0$,

$$\Phi_Q(t) - \rho_{0Q}(t) = e^{-\Gamma t}, \tag{3.23}$$

where

$$\Gamma = \lambda q c_Q F/(1 - \lambda - \lambda q f_{00}), \tag{3.24}$$

c_Q is the concentration of impurities, and we have used Eqs. (2.10), (2.21b), (3.10), and (3.11). For the one-trap problem with $\lambda = 1$, by the same reasoning that led to Eq. (2.20), we take $FQ_T = -\Gamma$. Then

$$\Phi_T(t) - \rho_{0T}(t) = \frac{1}{\Gamma(s_0 - s_1)}\left[(\Gamma - Fs_1)s_0 e^{Fs_1 t} - (\Gamma - Fs_0)s_1 e^{Fs_0 t}\right], \tag{3.25}$$

where s_0 and s_1 are given by Eq. (3.10) with $\lambda = 1$, $Q = -\Gamma/F$, and $N = 1/c_T$; c_T is the concentration of traps. Equation (3.25) is correct to $o(c_T)$. In the presence of both traps and impurities the time dependence of the host excitation is then

$$P_H(t) = e^{-(\Gamma + \tau_H^{-1})t}\left[\Phi_T(t) - \rho_{0T}(t)\right]. \tag{3.26}$$

The observable time dependences are those of the impurities $P_Q(t)$ and of the host + traps $P_H(t) + P_T(t)$. These are obtained from

$$\frac{d}{dt}\{e^{t/\tau_H}[P_H(t) + P_T(t)]\} = -\Gamma e^{t/\tau_H} P_H(t) \tag{3.27}$$

and

$$\frac{d}{dt}[e^{t/\tau_H} P_Q(t)] = e^{t/\tau_H}[\Gamma P_H(t) - QFP_Q(t)], \tag{3.28}$$

whose solutions are

$$P_H(t) + P_T(t) = e^{-t/\tau_H}\left(1 - \Gamma\int_0^t e^{t'/\tau_H} P_H(t')dt'\right) \tag{3.29}$$

and

$$P_Q(t) = e^{-t/\tau_Q}\Gamma\int_0^t e^{t'/\tau_Q} P_H(t')dt'. \tag{3.30}$$

For $\mu_T = 0$, the results given here are implicit in earlier work.[1] In real systems it is likely that $\mu_T > 0$. If there is a distribution of μ_T, one can use the CPA to treat the distribution. Since, in general, such a distribution is temperature dependent, one can, in principle, include this temperature dependence in the excitation quenching kinetics.

In a system with reversible impurities ($\mu_Q > 0$), the decay of $P_Q(t)$ is strongly dependent on the impurity concentration c_Q, particularly for low c_Q. The dependence is strongest with an impurity initial condition [see Eq. (3.11) and the discussion that follows it].

C. Anisotropic Lattices

We have limited our general formalism to isotropic lattices for simplicity. However, energy transfer rates in organic solids are anisotropic. We discuss here the influence of a simple but realistic anisotropy on the kinetics of nondisruptive quenching.

With the uniform initial condition, Eq. (2.16) becomes

$$FM_0 = (N/Q) - Nf_{00}. \tag{3.31}$$

Since there is only a single nondisruptive quencher, Eq. (3.31) is valid for any lattice with the appropriate f_{00}. We treat the case of a simple cubic lattice with only nearest-neighbor interactions modified so that the nearest neighbor energy transfer rates along the a, b, and c directions are F, F, and βF, respectively, where $0 \le \beta \le 1$.

In Eq. (3.6),

$$f(s) = -\frac{1}{N}\sum_{\mathbf{k}\neq 0}\left\{s + 4\left[\sin^2\left(\frac{\pi k_1}{N^{1/3}}\right) + \sin^2\left(\frac{\pi k_2}{N^{1/3}}\right) + \beta\sin^2\left(\frac{\pi k_3}{N^{1/3}}\right)\right]\right\}^{-1}, \tag{3.32}$$

where $\mathbf{k} = (k_1, k_2, k_3)$. For large N the sums in Eq. (3.32) can be approximated by a triple integral. One integration can easily be done to give, for $\beta \gg 1/N$,

$$f(s) = -\left(\frac{2}{\pi}\right)^2 \iint_0^{\pi/2} \frac{dx\,dy}{[s + 4(\sin^2 x + \sin^2 y)]^{1/2}[s + 4(\sin^2 x + \sin^2 y) + 4\beta]^{1/2}}. \tag{3.33}$$

For $|s| \lesssim N^{-1} \ll 1$, the sines in Eq. (3.33) can be approximated by their arguments. Then

$$f(s) \simeq -\frac{2}{\pi}\int_0^R \frac{r\,dr}{(s + 4r^2)^{1/2}(s + 4\beta + 4r^2)^{1/2}}, \tag{3.34}$$

where $x = r\cos\theta$, $y = r\sin\theta$, the original square region of integration is replaced by a quarter circle of radius R, and we have performed the θ integration. Since the effect of the approximations we have made is merely to add to $f(s)$ the quantity $C(R) + o(s)$ where $C(R)$ is independent of s, R is chosen to give the correct value of $\lim_{s\to 0} f(s)$ at $\beta = 1$ (we ignore the slight β dependence of R since it is immaterial in the results). Equation (3.34) gives

$$f(s) \simeq f_{00}(\beta = 1) + (1/2\pi)\ln(\tfrac{1}{2}\sqrt{s} + \tfrac{1}{2}\sqrt{s + 4\beta}). \tag{3.35}$$

The f_{00} to be used in Eq. (3.31) for the anisotropic simple cubic lattice is then

$$f_{00}(\beta) = -\tfrac{1}{6}c_1 + (1/4\pi)\ln\beta, \tag{3.36}$$

with the c_1 for the SC lattice given in Table I. Thus for $\beta \gg 1/N$ the zero mode is dominant and the zero moment retains its linear dependence on concentration. In other words, unless β becomes as small as $o(1/N)$ the anisotropic lattice behaves kinetically exactly as the isotropic lattice, although the coefficient of the concentration in FM_0 is larger than in the isotropic case.

As β approaches $o(1/N)$, features characteristic of two-dimensional kinetics begin to appear. For $\beta \ll 1/N$ the lattice behaves kinetically as a two-dimensional lattice.

The analysis of an anisotropic square lattice, having energy transfer rates F and βF, is quite similar to that of the cubic lattice. The kinetics of the lattice remains essentially two-dimensional until β approaches $o(1/N \log N)$. For $\beta \ll 1/N \log N$ the behavior is one-dimensional.

4. CONCLUSIONS

In this paper we have examined many aspects of incoherent exciton quenching. The results provide a fairly comprehensive picture of the quenching kinetics. While we have generalized some of our earlier one-dimensional results,[5] the emphasis has been on two- and three-dimensional lattices. In brief:

(1) The concept of a quencher has been generalized to include three physically significant parameters μ, λ, and Q. In the usual random-walk treatments[12,15–17] $\mu = 0, \lambda = 1$, and $Q = \infty$.

(2) The quantity qFM_0 given here is a generalization of earlier mean first passage time results.[5,16] In all cases of short-range quenching qFM_0 consists of a mean time to reach the nearest neighbors of the quencher plus a mean time to be quenched from those neighbors. As for the earlier results, qFM_0 has a dependence on quencher concentration that is characteristic solely of the number of dimensions.

(3) Under uniform initial excitation, except in one dimension or with slow back-transfer, the lowest eigenmode of the master equation dominates, i.e., except for correction terms that in three dimensions are of $o(c^{2/3})$ the excitation decay is a single exponential whose rate constant $\Gamma_0 = 1/M_0$.

(4) Whenever there is slow back-transfer from the quencher (i.e., when $F|Q| \sim \Gamma_0, F_\mu \sim \Gamma_0$ and $F_\lambda \gg \Gamma_0$), the detailed time dependences of the host and of the quencher change considerably from what they are when there is no back transfer.

(5) In three dimensions with a uniform initial condition the kinetics with random quencher locations differ from those for periodic locations only by $o(c^{1/2})$. In two dimensions, zero mode dominance is equally strong for random and periodic locations but the random M_0 is greater than the periodic M_0 by about 25% for $c \sim 10^{-4} - 10^{-6}$.

(6) In the presence of two types of defects, host traps of small $\mu > 0$ and impurity quenchers, the time dependences of the host and quencher are significantly altered just as for slow back-transfer from the quencher. In general, μ is a random variable whose distribution is temperature dependent.

(7) A three-dimensional lattice in which the nearest-neighbor energy transfer rate along one lattice direction (βF) is much less than the other two rates (F) still behaves kinetically as an isotropic three-dimensional lattice as long as $\beta \gg c$. An analogous conclusion holds for an anisotropic two-dimensional lattice.

Except for purely nondisruptive quenching we have not explicitly considered extended interactions in this paper. However, it is not difficult to deduce some of the principal qualitative features of quenching kinetics with extended interactions. For interactions that diminish exponentially with separation between all pairs of molecules (host as well as impurities) there is no qualitative change in the quenching kinetics in three dimensions.[18] If host–host interactions remain relatively short range while the range of the host-impurity interaction increases, no change occurs in any dimension until the range of the latter interaction becomes exceedingly long.[19]

---

* Contribution No. 4 from the Oak Ridge Summer Institute of Theoretical Biophysics.

† Operated by the U.S. Atomic Energy Commission under contract with the Union Carbide Corporation.

‡ Supported in part by NSF Contract No. GP 28257X, The Oak Ridge Associated Universities, and the Oak Ridge National Laboratory.

1 J. B. Birks, J. Phys. B3, 1704 (1970).
2 G. Hoch and R. S. Knox, in *Photophysiology*, edited by A. Giese (Academic, New York, 1968), Vol. 3.
3 R. B. Fox, T. R. Price, and R. F. Cozzens, J. Chem. Phys. **54**, 79 (1971); J. Eisinger and A. A. Lamola, in *Excited States of Proteins and Nucleic Acids*, edited by R. F. Steiner and I. Weinryb (Plenum, New York, 1971); J. W. Longworth, in *Excited States of Proteins and Nucleic Acids*.
4 See, e.g., Th. Förster, Ann. Physik **2**, 55 (1948); R. M. Pearlstein, Brookhaven Symp. Biol. **19**, 8 (1967); E. W. Montroll, Energetics Met. Phenomena **3**, 127 (1967).
5 K. Lakatos-Lindenberg, R. P. Hemenger, and R. M. Pearlstein, J. Chem. Phys. (to be published).
6 D. Bedeaux, K. Lakatos-Lindenberg, and K. E. Shuler, J. Math. Phys. **12**, 2116 (1971).
7 R. M. Pearlstein, K. Lakatos-Lindenberg, and R. P. Hemenger, Phys. Rev. Letters **27**, 1509 (1971).
8 C. Hélène and J. W. Longworth (in preparation).
9 P. Soven, Phys. Rev. **156**, 809 (1967); B. Velický, S. Kirkpatrick, and H. Ehrenreich, *ibid.* **175**, 747 (1968).
10 R. M. Pearlstein, J. Chem. Phys. **56**, 2431 (1972).
11 Equation (2.10) is equivalent to the quenching theorem of Ref. 10.
12 E. W. Montroll, Proc. Symp. Appl. Math. **16**, 193 (1964).
13 R. C. Powell, Phys. Rev. B **4**, 628 (1971).
14 Z. Bay and R. M. Pearlstein, Proc. Natl. Acad. Sci. (U.S.) **50**, 1071 (1963).
15 E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
16 E. W. Montroll, J. Math. Phys. **10**, 753 (1969).
17 H. B. Rosenstock, Phys. Rev. **187**, 1166 (1969).
18 K. Lakatos-Lindenberg and K. E. Shuler (in preparation).
19 K. Lakatos-Lindenberg, R. P. Hemenger, and R. M. Pearlstein (in preparation).

On the Mathematical Characterization of the Breaking of Hadronic Internal Symmetry

Lorenzo Abellanas*
Institut des Hautes Etudes Scientifiques, 91-Bures-Sur-Yvette, France
(Received 24 January 1972)

The adjoint action of a semisimple complex Lie algebra $\mathcal{G}$ singles out some singular directions with special properties. The directions of breaking of the hadronic internal symmetry groups, e.g., $SU(3)$ or $SU(3) \times SU(3)$, by the electromagnetic, semileptonic, and nonleptonic weak interactions, are of this type. They are Hermitian (Q, Y) or nilpotent (Cabibbo direction) elements in the corresponding complexified Lie algebra. Their mathematical characterization is provided by the adjoint action of $\mathcal{G}$. In particular, they satisfy a generalized extremal property for every $\mathcal{G}$-invariant polynomial.

INTRODUCTION

The hadron symmetry groups G, e.g., $SU(3)$ or $SU(3) \times SU(3)$), are broken by all fundamental interactions. They single out, however, well-defined directions of the Lie algebra of G. In the case of $SU(3)$ [we restrict our explicit statements to this, but all is easily extended to $SU(3) \times SU(3)$], these directions of breaking are defined by the following vectors in the customary Gell-Mann basis:

semistrong interactions: $Y = \lambda_8$ (hypercharge);

electromagnetic interactions:

$$Q = -\tfrac{1}{2}(\lambda_8 + \sqrt{3}\,\lambda_3) \text{ (electric charge);}$$

weak semileptonic interactions:

$$C_\pm = \tfrac{1}{2}(\lambda_1 \pm i\lambda_2)\cos\theta + \tfrac{1}{2}(\lambda_4 \pm i\lambda_5)\sin\theta = c_1 \pm ic_2 \text{ (Cabibbo directions);}$$

weak nonleptonic interactions:

$$\begin{aligned} Z &= c_1 \vee c_1 = c_2 \vee c_2 \\ &= \tfrac{1}{2}(3\cos^2\theta - 1)\lambda_8 \\ &\quad + \tfrac{1}{2}\sqrt{3}\sin^2\theta\,\lambda_3 \\ &\quad + \sqrt{3}\sin\theta\cos\theta\,\lambda_6 \text{ ("weak hypercharge").} \end{aligned}$$

Some attention has been recently devoted to the mathematical characterization of these directions. In particular, Michel and Radicati[1] have investigated this problem on the basis of a theorem concerning differentiable actions of compact groups on manifolds[2]. This theorem shows the existence of two four-dimensional critical orbits with the remarkable property that every smooth G-invariant function has an extremum on them. Since Y, Q, Z belong to these two critical orbits, the extremal property applies to them. The physical implications of this fact have been discussed in Ref. 1. If one adds to G the charge conjugation, then also c_1, c_2 become critical directions, for their stabilizers (= little groups) contains twice as many elements as those of almost all other elements of the Lie algebra of $SU(3)$.

In this paper we propose a different approach to the mathematical characterization of the directions of breaking of the hadron internal symmetry, based on the general theory of semisimple complex Lie algebras. Indeed, since the weak currents are electrically charged, the Cabibbo directions have to be complex linear combinations of the Hermitian Gell-Mann matrices. They are even nilpotent elements of $sl(3, C)$, the complexified Lie algebra of $SU(3)$. We must emphasize that the object to be complexified is the Lie algebra itself, rather than its underlying vector space. In complexifying the Lie algebra, we are led to consider at the same time the adjoint action of $SL(3, C)$. Note that the symmetric algebra ($\vee$-algebra) of $SU(3)$ was already complexified in Ref. 1.

In order to justify the use of complex Lie algebras, we could remark that in $SU(2)$ itself, when one wants to study the $SU(2)$ breaking, non-Hermitian operators turn out to be essential. Indeed, the electromagnetic current singles out the generator T_3 and the weak currents the generators T_+, T_-. These three elements generate the noncompact $SL(2, \mathbb{R})$ or $SL(2, C)$ Lie algebras.

To make this paper as self contained as possible, we begin by a brief review of some important results about semisimple complex Lie algebras and their adjoint actions. In Sec. 1 we give the essential lemmas on orbit structure that we shall need. Section 2 deals with invariant polynomials. The main result is due to Konstant.[3] General references for these topics are Refs. 3, 4. As a help for understanding these two first sections, the simplest example of $SL(2, C)$ is presented in Sec. 3.

In Sec. 4A *singular* orbits are defined. In Sec. 4B we investigate the structure of the set n of *nilpotent*) elements of the Lie algebras A_l. This will show the existence of *nilpotent* singular orbits in the A_l for the adjoint action.

Section 5 contains an explicit analysis for $SL(3, C)$, on which are based the conclusion of this paper: We show that there exist an infinity of semisimple singular orbits homothetic to any *one* of them and *one* nilpotent singular orbit. We find also that, when restricted to the real (Hermitian) Lie algebra of $SU(3)$, the definition of singular orbits coincides with the definition of critical orbits as given in Ref. 2.

The physical conclusion is that Y, Q, Z belong to semisimple singular orbits and $c_\pm$ to the nilpotent singular orbit. This implies the following properties for these elements:

(i) They satisfy the equation $x \vee x = \mu x$ with $\mu \neq 0$ (resp. $\mu = 0$) for Y, Q, Z (resp. $c_\pm$).

(ii) For any invariant polynomial Φ defined on the Lie algebra of $SL(3, C)$, one has $d\Phi_x = \lambda x$. This is a generalized version of the extremal property proved in Refs. 1, 2.

(iii) The stabilizers of all these elements are maximal among all the stabilizers which appear in the adjoint action of $SL(3, C)$.

1. ORBIT STRUCTURE OF THE ADJOINT REPRESENTATION OF A SEMISIMPLE COMPLEX LIE ALGEBRA

Let $\mathcal{G}$ be a semisimple complex Lie algebra (s.c.L.a.) of dimension n. The adjoint group G consists of all $\exp(\operatorname{ad} x)$, $x \in \mathcal{G}$. Of course its Lie algebra is isomorphic to $\mathcal{G}$. The group G acts on $\mathcal{G}$ by: $x \to axa^{-1}$,

$x \in \mathcal{G}, a \in G$.

The differential at e of these inner automorphisms yields a set of linear automorphisms of $\mathcal{G}$, which is called the *adjoint representation* of $\mathcal{G}$.

As usual, we will denote G_x the stabilizer or little group of $x \in \mathcal{G}$, and $\mathcal{O}_x = \{axa^{-1} \mid a \in G\}$ the orbit of x. We will speak indistinctly of $\mathcal{G}_x$ or G_x as the stabilizer of the point x. By $\mathcal{G}_x$ we mean the Lie algebra of G_x, i.e., the centralizer of x in $\mathcal{G}$.

Definition 1: $x \in \mathcal{G}$ is said to be *semisimple* (resp. *nilpotent*) if ad x is a diagonalizable (resp. nilpotent) endomorphism of $\mathcal{G}$. It will be noted $x \in \mathcal{S}$ ($x \in \mathfrak{N}$).

That both of these concepts are fundamental in the analysis of s.c.L.a., is shown by the following[3]:

Lemma 1: For each $x \in \mathcal{G}$ we can find $s \in \mathcal{S}$, $n \in \mathfrak{N} \ni x = s + n$ and $[s, n] = 0$. These s, n will be referred to as the semisiple and nilpotent parts of x. (They are unique.)

It is a well-known fact that for real differentiable ($\equiv C^\infty$) G spaces, with G a compact Lie group, every orbit is closed. This property does not remain true for arbitrary situations. However, in the case we deal with here, we have

Lemma 2. Let $\mathcal{G}$ be an s.c.L.a. Then $\mathcal{O}_x$ is *closed* in $\mathcal{G}$ iff x is semisimple. Since the semisimple elements are dense in $\mathcal{G}$ (see below), it follows that the union of all closed orbits is dense in $\mathcal{G}$.

Notice that the only $x \in \mathcal{S} \cap \mathfrak{N}$ is $x = 0$. So the only closed nilpotent orbit is $\mathcal{O}_0 = \{0\}$.

By looking at the equality $\dim\mathcal{G} = n = \dim\mathcal{O}_x + \dim\mathcal{G}_x$, we see that the orbits can be indexed by the dimension (over $\mathbb{C}$) of their centralizer. There is a very important *partial ordering:* $\mathcal{G}_x \leq \mathcal{G}_y$ iff $\dim\mathcal{G}_x \leq \dim\mathcal{G}_y$ and $\mathcal{G}_x$ is conjugated in $\mathcal{G}$ to a subalgebra of $\mathcal{G}_y$. Under this ordering, the orbits which appear in the adjoint action of $\mathcal{G}$ become ordered in such a manner that $\mathcal{O}_x \leq \mathcal{O}_y \Longleftrightarrow \mathcal{G}_y \leq \mathcal{G}_x$.

For a compact G, a theorem by Montgomery asserts that given a differentiable G-space X, the set M of *maximal orbits* is dense in X and moreover $\dim(X - M) \leq \dim X - 2$. (See Ref. 5).

When one drops the compacity of G, this fails to be true in general. Again semisimplicity is strong enough to preserve it, as we are going to state in a moment. Before doing it, we recall some definitions.

Let $\mathcal{G}$ be an s.c.L.a. Then we consider the characteristic equation for ad x, namely $\det(\beta\text{-ad } x) = \sum_0^n F_i(x)\beta^i = 0$. The rank l of $\mathcal{G}$ is defined as the smallest integer such that $F_l(x) \not\equiv 0$. One easily sees that $\dim\mathcal{G}_x \geq l$. An element $x \in \mathcal{G}$ is called *regular* if $F_l(x) \neq 0$. Obviously the set $\mathcal{R}$ of regular elements is open and dense in $\mathcal{G}$ (just consider that it is the complement of the null set of an analytic function $\not\equiv 0$). One can show that $\mathcal{R} \subset \mathcal{S}$ (see Ref. 4).

Given $x \in \mathcal{G}$, the $\mathcal{G}_x$ contains always an l-dimensional Abelian subalgebra of $\mathcal{G}$.[3] For instance, if $x \in \mathcal{R}$, $\mathcal{G}_x$ is a Cartan subalgebra of $\mathcal{G}$. This is an important example of minimal stabilizer. In order to find a criterion for elements lying in maximal orbits, we need one more definition.

Definition 2: Let us denote by $\mathfrak{M}$ the set $\{x \in \mathcal{G} \mid \dim\mathcal{G}_x = l \text{ (minimal!)}\}$. An element $x \in \mathfrak{N}$ is said to be *principal nilpotent* if $x \in \mathfrak{N} \cap \mathfrak{M}$.

The following lemmas can be found in Ref. 3.

Lemma 3: Let x, s, n be as in Lemma 1 above. Then

$$x \in \mathfrak{M} \Longleftrightarrow n \text{ is principal nilpotent in } \mathcal{G}_s.$$

Take for instance $x = s \in \mathcal{S}$ such that $\mathcal{G}_x$ is a Cartan subalgebra. Then $n = 0$ is principal in $\mathcal{G}_x$. On the other hand $x = n \in \mathfrak{N}$ is in $\mathfrak{M}$ iff n is principal nilpotent in $\mathcal{G}$.

Lemma 4: $\dim\mathcal{O}_x = \text{even}, \forall\, x \in \mathcal{G}$.

Proof: Let us define $K_x(y, z) = K([z, y], x)$, where K stands for the nondegenerate invariant Killing form on $\mathcal{G}$. Then K_x is also an alternating bilinear form on $\mathcal{G}$. Moreover, $K_x(y, z) = 0, \forall\, z \in \mathcal{G} \Longleftrightarrow K([z, y], x) = K(z, [y, x]) = 0, \forall\, z \in \mathcal{G} \Longleftrightarrow y \in \mathcal{G}_x$. Thus K_x is nondegenerate on $\mathcal{G}/\mathcal{G}_x$, so K_x must have an even-dimensional carrier space, i.e., $\dim\mathcal{G}/\mathcal{G}_x = \dim\mathcal{O}_x = \text{even}$. QED

Corollary: If $x \notin \mathfrak{M}$, then $\dim\mathcal{O}_x \geq l + 2$. (Compare this with Montgomery's results!).

We know (from Lemma 1) that the existence of nonclosed orbits in $\mathcal{G}$ is essentially due to the nilpotent elements. The orbit structure of $\mathfrak{N}$ is explained by the following:

Lemma 5: (a) There are only a finite number of orbits in $\mathfrak{N}$. (b) There is one $(n - l)$-dimensional orbit which is dense in $\mathfrak{N}$

Notice in particular that $\dim\mathfrak{N} = n - l$. The dense orbit consists of all principal nilpotent elements.

2. THE RING OF INVARIANT POLYNOMIALS

The orbit space $\mathcal{G}/G$ can be parametrized by a complete set of algebraically independent invariant polynomials. It is well known[6] that the ring of G-invariant polynomials $\mathcal{J}$ is generated by l homogeneous polynomials $c_1, \ldots, c_l$ such that $\sum_1^l \deg c_i = \frac{1}{2}(n + l)$, where $n = \dim\mathcal{G}$ and $l = $ rank of $\mathcal{G}$.

In particular, for A_l one possible choice is the following: $c_i(x) = \operatorname{tr}(x)^{i+1}$, where x stands for the evident $(l + 1) \times (l + 1)$ representation of a generic element of A_l. (See e.g. 7.)

From now on, we will assume the set $(c_1, \ldots, c_l)$ to be indexed so that $i \leq j \Longrightarrow \deg c_1 \leq \deg c_j$. An explicit parametrization for $\mathcal{G}/G$ is given by the map $\varphi: \mathcal{G}/G \to \mathbb{C}^l$ defined as $\varphi(\mathcal{O}_x) = (c_1(x, \ldots, c_l(x))$. It can be shown[3] that when restricted to $\mathcal{S} \cup \mathfrak{M}$, the map φ is a bijection. However, this is not true for the restriction of φ to $\mathfrak{N}$. Indeed the element $x = s + n$ is conjugated to $x' = s + \lambda n$ ($\lambda \neq 0$), and hence the value at x of any $\pi \in \mathcal{J}$ depends only on the semisimple part of x. More precisely we have[3]:

Lemma 6: The cone $\langle x \in \mathcal{G} \mid \pi(x) = 0, \forall\, \pi \in \mathcal{J}$ without constant term$\rangle$ coincides with $\mathfrak{N}$.

We state a basic theorem[3] which characterizes $\mathfrak{M}$.

Theorem 1: Let $\mathcal{G}$ be a s.c.L.a. and $x \in \mathcal{G}$. Then the following conditions are equivalent:

(a) $x \in \mathfrak{M}$;

(b) $dc_i(x)$ $(i = 1, 2, \dots, l)$ are linearly independent at x.

3. AN EXAMPLE: $sl(2, C) = A_1$

We have $n - l = 2$, so we conclude (from Lemma 4 above) that $x \notin \mathfrak{M} \Rightarrow \dim \mathcal{G}_x = 3 \Rightarrow \mathcal{G}_x = \mathcal{G} = sl(2, C)$. This is very easily checked by classifying the orbits of the adjoint action according to Jordan canonical forms. There are only three types:

$$\overset{(1)}{\begin{pmatrix} \lambda & \\ & -\lambda \end{pmatrix}} \quad \overset{(2)}{\begin{pmatrix} 0 & \lambda \\ 0 & 0 \end{pmatrix}} \quad \overset{(3)}{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}}.$$

These are the three strata[8] in $\mathbb{C}^3$, the underlying vector space of $sl(2, C)$. We list some properties:

(i) The stabilizers are: $\mathcal{G}_{(1)}$ = Cartan subalgebra, $\mathcal{G}_{(2)} = U_c(1)$, $\mathcal{G}_{(3)} = sl(2, \mathrm{C})$. Therefore we see that $\mathfrak{M} = (1) \cup (2)$ and

$$x = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \notin \mathfrak{M}$$

has $\mathcal{G}_x = \mathcal{G}$.

(ii) $\mathcal{S} = (1) \cup (3)$, $\mathfrak{N} = (2) \cup (3)$.

(iii) There are an infinity of orbits in (1), one for each value of $\lambda \in \mathbb{C}$.

(iv) There are only two nilpotent orbits. In fact that (2) consists of a unique orbit follows from the fact that

$$x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } x' = \begin{pmatrix} 0 & \beta \\ 0 & 0 \end{pmatrix}$$

are conjugate: $x' = gxg^{-1}$, where

$$g = \begin{pmatrix} \beta^{1/2} & \\ & \beta^{-1/2} \end{pmatrix}.$$

(v) (1) is dense in $\mathcal{G}$ and (2) is dense in $\mathfrak{N}$.

(vi) The map φ (see Sec. 2) does not separate the orbits in $\mathfrak{N}$. Take

$$a = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } b = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then $a, b \in \mathfrak{N}$ and $\mathcal{O}_a \neq \mathcal{O}_b$. But $\varphi(\mathcal{O}_a) = \varphi(\mathcal{O}_b) = 0 \in \mathbb{C}$.

Compare these trivial facts with the general setting in Secs. 1 and 2. In a basis $\{\sigma_3, (\sigma_1 \pm i\sigma_2)/\sqrt{2}\}$, we have $\sigma_3 \in (1)$ and $(\sigma_1 \pm i\sigma_2/\sqrt{2}) \in (2)$ (where σ_j are the usual Pauli matrices).

4. SINGULAR ORBITS

A. The Rank Function

A real-valued function f: $X \to \mathbb{R}$ is said to be lower semicontinuous (l.s.) at $x \in X$ if and only if $\forall\, \alpha < f(x)$, $f^{-1}(\alpha, +\infty]$ is a neighborhood of x in X. If f is l.s. on X, then $\{x \in X \mid f(x) \leq \beta\}$ is closed $\forall\, \beta \in \mathbf{R}$.

Now consider a finite-dimensional vector space E over $\mathbb{C}$, and let $L(E)$ be the set of all bounded linear operators on E. For $A \in L(E)$ we define $\operatorname{rank}(A) = \dim\{Av \mid v \in E\}$. This function ("rank") is lower semicontinuous (in the strong topology).[9]

We are interested in the case where $E = \mathcal{G}$ is a s.c.L.a. and $A = d\varphi_x \in L(E)$. The map $d\varphi_x = ((dc_1)_x, \dots, (dc_l)_x)$ is the differential of φ (Sec. 2) evaluated at $x \in \mathcal{G}$. [We identify $T_x(\mathcal{G})$, the tangent space at x, to $\mathcal{G}$ itself.]

The Killing form K on $\mathcal{G}$ permits to define the dual vector grad f_x as the unique vector in $\mathcal{G}$ which satisfies $K(\operatorname{grad} f_x, v)_x = df_x(v), \forall\, v \in T_x\mathcal{G}$.

Let $\rho(x) = \operatorname{rank} d\varphi_x$, and put $\mu = \inf\{\rho(x) \mid x \in \mathcal{G} - (0)\}$.

Definition 3: We will say that the orbit $\mathcal{O}_x$ is *singular* if and only if $\rho(x) = \mu$.

Since ρ is l.s., the set of $\{x \in \mathcal{G} \mid \mathcal{O}_x$ is singular$\}$ is closed in $\mathcal{G} - (0)$. One can look now at Theorem 1 as giving a characterization of those elements $x \in \mathcal{G}$ where $\rho(x) = l$ (maximal!). It can be reformulated as follows:

Theorem 2: Let $\mathcal{G}$ be an s.c.L.a. Then $x \in \mathfrak{M} \Longleftrightarrow \rho(x) = l$.

Now let $\Phi(c_1, \dots, c_l)$ be an invariant polynomial.

Then $d\Phi_x = \sum_1^l \Phi_j (dc_j)_x$, where $\Phi_j = \partial\Phi/\partial c_j$. Theorem 1 implies strong restrictions on $d\Phi$ at least in those situations where l is small. For physical application the most relevant case is $l = 2$.

B. The orbit structure of $\mathfrak{N}$

Remember that we have indexed $(c_1, \dots, c_l)$ so that $i \leq j \Rightarrow \deg c_i \leq \deg c_j$. With this notation, a theorem of Varadarajan[10] asserts that $(dc_l)_x = 0$, $\forall x \in \mathfrak{N}$. (This is a more explicit result than theorem 1). We would like to point out a slight improvement of this theorem for the algebras A_l, which may be useful in order to understand the structure of $\mathfrak{N}$.

Consider the set $\mathfrak{N}_i = \{x \in \mathfrak{N} \mid x^{i+1} = 0 \neq x^i\}$. Since $x \in \mathfrak{N} \Rightarrow x^k = 0$ for $k \geq l + 1$, we see that $\mathfrak{N} = \cup_{i=0}^l \mathfrak{N}_i$. The canonical Jordan form for $x \in \mathfrak{N}_i$ is a $(l + 1) \times (l + 1)$ matrix with i 1's on the diagonal $x_{j,j+1}$, i.e., $x_{kl} = \sum_{j=1}^i \delta_{k,j}\delta_{l,j+1}$. Therefore $i \geq j \Rightarrow \dim \mathcal{G}_{\mathfrak{N}_i} \leq \dim \mathcal{G}_{\mathfrak{N}j}$.

Maximal stabilizers (up to the trivial $\mathfrak{N}_0 = \{0\}$ case) correspond to $\mathfrak{N}_1$. We obtain in this way a graded structure in $\mathfrak{N}$, which induces an ordering for the stabilizers.

Now consider the basis $c_i(x) = \operatorname{tr} x^{i+1}$ for $\mathcal{I}$ (Sec. 2). If $x \in \mathfrak{N}$, a trivial computation yields: $(dc_i)_x = (i + 1)x^i$. Hence $x \in \mathfrak{N}_j \Rightarrow (dc_k)_x = 0$, $k > j \Rightarrow \rho(x) = j$.

This shows that according to the previous definition,

the orbits in $\cup_{i=2}^l \mathfrak{N}_i$ cannot be singular.

5. APPLICATION TO HADRONIC INTERNAL SYMMETRY

A. Complex Lie Algebra of $SU(3)$: $SL(3, C) = SU_c(3)$.

We denote it by $\mathcal{G}$ in this paragraph. Its underlying complex vector space is $\mathbb{C}^8$, whose elements are 3×3 traceless complex matrices. The invariant bilinear form reads

$$(x, y) = \tfrac{1}{2}\mathrm{tr}\, xy.$$

From $\dim \mathrm{Hom}\,(\mathcal{E}\otimes\mathcal{E}, \mathcal{E})^{\mathcal{G}} = 2$, where $\mathcal{E}$ stands for the adjoint representation of $\mathcal{G}$, we conclude the existence of two independent invariant algebras. Following Ref. 1 we take them to be

Lie algebra: $x_\wedge y \equiv -\tfrac{1}{2}i[x, y]$

Symmetric algebra: $x_\vee y \equiv \tfrac{1}{2}\sqrt{3}\,(xy + yx) - (2/\sqrt{3})\,(x,y)$.

It is easy to see that both definitions yield bilinear maps $\mathcal{G} \times \mathcal{G} \to \mathcal{G}$.

Conventions: We introduce the auxiliary ($\notin \mathcal{G}$) matrices $1, d_1, d_2, d_3$ with $(d_i)_{jk} = \delta_{jk}\delta_{ij}$. The elements of $\mathcal{G}$ will be denoted by:

diagonal matrices:

$$\left.\begin{aligned} q_i &= (1/\sqrt{3})\,(1 - 3d_i) \\ r_i &= (1/\sqrt{3})\,(q_{i-1} - q_{i+1}) \end{aligned}\right\} \sum_1^3 q_i = \sum_1^3 r_i = 0$$

$$(i + 3 \equiv i); \tag{1}$$

nondiagonal matrices:

$$\begin{aligned} z_i &\ni (z_i)_{\alpha\beta} \equiv \sqrt{2}\,\eta_{i\alpha\beta} \equiv \sqrt{2}\,\delta_{j,i+1}\delta_{j,k-1}, \\ z_i^* &\ni (z_i^*) \equiv \sqrt{2}\,\eta_{i\beta\alpha}. \end{aligned} \tag{2}$$

We have, for instance,

$$q_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} -2 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \quad r_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 0 & & \\ & 1 & \\ & & -1 \end{pmatrix},$$

$$z_1 = \sqrt{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad z_1^* = \sqrt{2}\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

A possible choice for a basis of $\mathcal{G}$ is the following:

$$\{q_1,\ r_1, z_1, z_2, z_3, z_1^*, z_2^*, z_3^*\}.$$

Hermitian conjugation: $(x, y)^* = (x^*, y^*)$, $(x_\tau y)^* = x_\tau^* y^*$, where $\tau = \wedge, \vee$.

Lie algebra: a, b diagonal $\Longrightarrow a_\wedge b = 0$;

$$\begin{aligned} &a \wedge z_i = -i(a, r_i)z_i \xrightarrow{\ *\ } a \wedge z_i^* = i(a, r_i)z_i^*; \\ &z_i \wedge z_j = -(i/\sqrt{2})(\eta_{ijk} - \eta_{ikj})z_k^*, \quad z_i \wedge z_j^* = -i\delta_{ij}r_i. \end{aligned} \tag{3}$$

Symmetric algebra:

$$\left.\begin{aligned} r_{i\vee} r_j &= \delta_{ij}q_i + |\eta_{ijk} - \eta_{ikj}|q_k = -q_{i\,\vee}\,q_j \\ q_{i\vee} r_j &= \delta_{ij}r_i + |\eta_{ijk} - \eta_{ikj}|\,r_k \\ z_{i\vee} z_j &= (\sqrt{3}/2)\eta_{ijk} - \eta_{ikj}|z_k^* \\ z_{i\vee} z_j^* &= \delta_{ij}\,q_i \\ a_\vee z_i &= (q_i, a)z_i \end{aligned}\right\} \Longrightarrow \begin{aligned} r_{i\vee} r_i &= q_i \\ q_{i\vee} q_i &= -q_i, \\ q_{i\vee} r_i &= r_i, \\ z_{i\vee} z_i &= 0 = z_{i\vee}^* z_i^*. \end{aligned} \tag{4}$$

From (3) we see that $\pm r_i$ are the six roots of $\mathcal{G}$.

B. Orbit Structure on $\mathbb{C}^8$ Under the Adjoint Action of $SU_c(3)$

Let us write down all possible canonical forms for the elements in $\mathcal{G}$, like we did in the preceding section. We get six types:

$$\overset{\text{I}}{\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}} \overset{\text{II}}{\begin{pmatrix} \lambda & & \\ & \lambda & \\ & & -2\lambda \end{pmatrix}} \overset{\text{III}}{\begin{pmatrix} \lambda & 1 & \\ 0 & \lambda & \\ & & -2\lambda \end{pmatrix}} \overset{\text{IV}}{\begin{pmatrix} 0 & 1 & 0 \\ & 0 & 1 \\ & & 0 \end{pmatrix}} \overset{\text{V}}{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}} \overset{\text{VI}}{\begin{pmatrix} & & \\ & 0 & \\ & & \end{pmatrix}}$$

$\sum \lambda_i = 0, \quad \lambda_i \neq \lambda_j,\ i \neq j.$

Let us complete this classification with some relevant remarks concerning their elements and centralizers.

(I) This stratum is dense in $\mathbb{C}^8$. Their elements are regular, with $\mathcal{G}_I$ = Cartan subalgebra $U_c(1) \times U_c(1)$ of $\mathcal{G}$.

(II) A simple calculation shows that $x \in \text{II} \Longrightarrow x_\vee x = \alpha x$ with $\alpha \in \mathbb{C}$, i.e., they are idempotent in the V algebra. In other words, the semisimple x with two identical eigenvalues satisfy $x_\vee x = \alpha x$. Moreover if

$$x = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & -2\lambda \end{pmatrix}$$

and $\lambda \neq 0$, then $\alpha \neq 0$.

The Hermitian elements of this type were called "generalized charges" in Ref. 1. Observe that elements in II are semisimple but $\notin \mathfrak{M}$. Indeed every matrix like

$$\begin{pmatrix} U & 0 \\ 0 & -\mathrm{tr}U \end{pmatrix} \text{ commutes with } \quad x = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & -2\lambda \end{pmatrix},$$

wherever $U \in U_c(2)$. Hence $\mathcal{G}_{\text{II}} \approx U_c(2) \Longrightarrow \dim \mathcal{G}_x = 4$ (we mean always complex dimension). Remember Lemma 4!

(III) These elements, linear combinations of II and V, are not diagonalizable. We have for each $x \in \text{III}$ a decomposition (Lemma 1) $x = s + n$, $s \in \text{II}$, $n \in \text{V}$. In fact one immediately sees that

$$x = \begin{pmatrix} \lambda & \alpha & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & -2\lambda \end{pmatrix} \Longrightarrow s = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & -2\lambda \end{pmatrix} \in \text{II} \quad \text{and}$$

$$n = \begin{pmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in V$$

TABLE I

	Characterization	Stabilizer	Dimension (Over $\mathbb{C}$) Stratum	|Orbit|	Number of orbits	Examples	
I	$\gamma^3 \neq \theta^2$	$U_c(1) \times U_c(1)$ Cartan	8	6	∞^2	regular, e.g., r_i	$\subset \mathcal{S}$
II	$\gamma^3 = \theta^2 \neq 0$ x semisimple	$U_c(2)$	5	4	∞	$\pm q_i$	$\subset \mathcal{S}$
III	$\gamma^3 = \theta^2 \neq 0$ x nondiagonalizable	$U_c(1) \times U_c(1)$	7	6	∞	$\alpha q_i + \beta z_i \; (\alpha\beta \neq 0)$	$\subset \mathcal{S} + \mathfrak{N}$
IV	$\gamma = \theta = 0$ $x^3 = 0 \neq x^2$	$U_c(1) \times U_c(1)$	6	6	1	$\alpha z_1 + \beta z_2 + \gamma z_3^* \; (\alpha\beta \neq 0)$	$\subset \mathfrak{N}$
V	$\gamma = \theta = 0$ $x^2 = 0 \neq x$	Solvable four-dimensional	4	4	1	$\alpha z_i + \beta z_j^* \; (i \neq j)$	$\subset \mathfrak{N}$
VI	$x = 0$	$SU_c(3)$	0	0	1	$x = 0$	$\subset \mathcal{S} \cap \mathfrak{N}$

The strata II, V contain all singular orbits. Of course, ∞^r means that the codimension of orbit in its stratum is r. The index c in $U_c(1), SU_c(3)$, etc., means "complex Lie algebra" of $U(1), SU(3)$, etc.

which trivially satisfy $[s,n] = 0$. Since $\mathcal{G}_{\rm II} = U_c(2)$ and from our example Sec. 2 we know[11] that $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is principal nilpotent in $U_c(2)$, Lemma 3 yields the conclusion III $\subset \mathfrak{M}$. It is interesting to emphasize that $\dim \mathcal{G}_{\rm III} = 2$, e.g., $\mathcal{G}_x = \{q_i, z_i\}$ for $x = \alpha q_i + \beta z_i$; but $\mathcal{G}_{\rm III}$ is *not* isomorphic to *Cartan subalgebras*. More generally it can be shown for any s.c.L.a. $\mathcal{G}$ that every maximal Abelian subalgebra containing only semisimple elements is a Cartan subalgebra. This was actually the case in (I). In addition, there exists many other maximal Abelian subalgebras which contain nilpotent ($\neq 0$) elements, so that they are not of Cartan type.[3,4]

(IV) We find again $\dim \mathcal{G}_x = 2$ for $x \in$ IV. As an example take $x = \alpha z_1 + \beta z_2$ $(\alpha\beta \neq 0)$. Then

$$x = \begin{pmatrix} 0 & \alpha & 0 \\ 0 & 0 & \beta \\ 0 & 0 & 0 \end{pmatrix} \Longrightarrow \mathcal{G}_x = \{\alpha z_1 + \beta z_2, z_3^*\},$$

an Abelian two-dimensional subalgebra of $\mathcal{G}$. Since it also contains nilpotent elements, $\mathcal{G}_x$ is not of Cartan type. Thus we have three strata of maximal orbits: $\mathfrak{M} = \mathrm{I} \cup \mathrm{III} \cup \mathrm{IV}$; the last one being contained in $\mathfrak{N}$. In fact we have $\mathrm{IV} = \mathfrak{M} \cap \mathfrak{N}$.

(V) These are the elements in $(\mathfrak{N} - \mathfrak{M}) - \{0\}$. As a typical one we consider for instance $x = z_1 \in V \Longrightarrow \mathcal{G}_{z_1} = \{z_1, q_1, z_2^*, z_3^*\}$. More generally $\mathcal{G}_{zi} = \{z_i, q_i, z_j^*, z_k^*\}$ with $(i,j,k) = (1,2,3)$ and cyclics. Up to now (and there are no more!) we have two types of four-dimensional stabilizers corresponding to strata II, V. Notice, however, a crucial difference between both of them, namely $\mathcal{G}_{\rm II}$ contained a semisimple part isomorphic to $SU_c(2)$, while on the contrary $\mathcal{G}_{\rm V}$ are *solvable* subalgebras.

It may be illuminating to say a few words about solvable subalgebras of $\mathcal{G}$. Remember that a Lie algebra L is said to be solvable iff $D^k L = 0$ for some k, where D^k is definined by $D^i L = [D^{i-1}L, D^{i-1}L]$ and $D^0 L = L$. The maximal solvable Lie subalgebras of a s.c.L.a. $\mathcal{G}$ are called Borel subalgebras of $\mathcal{G}$. Their dimension is $\frac{1}{2}(n + l)$ (see Ref. 4). Therefore in our case $\frac{1}{2}(n + l) = 5$. An example of Borel subalgebra is the following: $\{q_i, r_i, z_i, z_j^*, z_k^*\}$ with $(i,j,k) = (1,2,3)$ or cyclic permutations.

By the other hand, one can demonstrate that every solvable subalgebra of $\mathcal{G}$ is contained in a Borel subalgebra. We see that $\mathcal{G}_{\rm V}$ are contained too. Observe that the *Borel subalgebras* could not be stabilizers in the adjoint action of $\mathcal{G}$, because of Lemma 4. However, we find four-dimensional solvable stabilizers, and we remark that they are obviously maximal in the above defined partial ordering for stabilizers.

(VI) This is an uninteresting case. (It is often convenient to work in the manifold $\mathbb{C}^8 - \{0\}$). When we speak of maximal stabilizers, we are implicitly assuming them to be proper subalgebras of $\mathcal{G}$.

We are now in position to write a table for the different strata of the adjoint action. In order to characterize the orbits we need two ($l = 2$) independent invariant polynomials. We choose θ, γ defined by $\gamma(x) = (x,x)$ and $\theta(x) = (x_\vee x, x)$, $x \in \mathcal{G}$. (We take $\gamma \equiv \frac{1}{2} c_1$ and $\theta \equiv \frac{1}{2} c_2$ with c_1, c_2 as defined in Sec. 2A.) The characteristic equation for $x \in \mathcal{G}$ is simply

$$x^3 - (\tfrac{1}{2}\, \mathrm{tr} x^2) x - (\tfrac{1}{3}\, \mathrm{tr} x^3)\mathbb{1} = x^3 - \gamma(x) x - (2/3\sqrt{3})\theta(x)\mathbb{1} = 0.$$

This is a very intuitive way of convincing oneself that we need exactly two quantities and that θ, γ is a natural choice.

The classification, according to the previous paragraph, is given in next table.

Let us observe that it is because of the existence of only a finite number of strata that we are able to write explicitly their table. This is a nontrivial remark. Indeed we know that compacity (for both the group G and the manifold X) is sufficient to prove this fact. In general, when one drops compacity that property is lost. But in our present situation, once again semisimplicity provides a sufficient condition.

C. Properties of the Singular Orbits for the Adjoint Action of $SL(3, C)$

With the aid of the previous table, we are able to give a complete characterization for singular orbits. The situation can be summarized as follows:

Singular orbits for $SL(3,C)$: Let φ be the mapping defined in Sec. 2. For $x \in sl(3,C) - \{0\}$, the following properties are equivalent:

(a) $x \notin \mathfrak{M}$;

(b) $\mathcal{O}_x \in \mathrm{II} \cup \mathrm{V}$;

(c) $x \vee x = \mu x$ for some $\mu \in \mathbb{C}$;

(d) $d\Phi_x = \lambda x$ $(\lambda \in \mathbb{C})$ $\forall\, \Phi \in \mathcal{I}$;

(e) $\mathcal{O}_x$ is singular.

Proof: (a) $\Longleftrightarrow$ (b): Table I makes it evident.

(a) $\Longleftrightarrow$ (e): follows from Theorem 1 and the fact that rank of $sl(3,C) = 2$.

(e) $\Longleftrightarrow$ (d): The only nontrivial part is (e) $\Longrightarrow$ (d). Notice, however that $(dc_1)_x = 2x$, and hence $(dc_1)_x \neq 0$, and the result follows.

(c) $\Longleftrightarrow$ (b): This is a trivial calculation. One finds $\mu = 0$ for $x \in \mathrm{V}$ and $\mu \neq 0$ for $x \in \mathrm{II}$.

Thus we see that there are two types of singular orbits. Firstly, we find an infinity of them in II. They correspond to different values of c_1, and in fact, if one is only interested by singular directions (as it is the case), it suffices to look at one of them. Since these orbits are semisimple, this is to be compared with the real (compact) case, investigated by Michel and Radicati.[1] They concluded that $d\Phi_x = 0$ for every (smooth) invariant function Φ on the sphere $S^7 \subset \mathbb{R}^8$.

But this is equivalent to say that $d\Phi_x$ is proportional to x, in the framework of $\mathbb{R}^8$. Thus we have rediscovered that result for semisimple elements.

Secondly, there is a unique singular orbit in V. As in case II, the stabilizer has dimension four. But if $x \in \mathrm{V}$ we know that x is conjugated to λx $(\lambda \neq 0)$, i.e., one of the four remaining generators moves x on λx, and hence it conserves the direction of x. In other words, when speaking about stabilizers of directions in the Lie algebra, the greatest one (five-dimensional) corresponds to the direction of elements $x \in \mathbf{V}$.

ACKNOWLEDGMENTS

I am indebted to Professor L. Michel for many enlightening discussions, criticisms, and helpful suggestions. I also wish to thank Professor L. Motchane and Professor L. Michel for their warm hospitality at the Institut des Hautes Etudes Scientifiques. This work has been supported by GIFT (Spain).

---

* On leave of absence from the University of Madrid, Spain. Permanent address: Facultad de Ciencias, Dept. Física Teórica, Universidad de Madrid, Madrid 3.

1 L. Michel and L.A. Radicati, *Symmetry Principles at High Energy. Fifth Coral Cables Conference*, p. 19, Benjamin 1968; L. Michel, *Lectures in Theoretical Physics* (U. of Colorado Press, Boulder, 1968), Vol. 10, p. 263; L. Michel, *Lectures Notes in Physics*, no. 6, 1969 Batelle Rencontres, Seattle (Springer-Verlag, Berlin, 1970); L. Michel and L.A. Radicati, Ann. Phys. (N.Y.) **66** 758 (1971).

2 L. Michel, Compt. Rend. **272**, 433 (1971).

3 B. Konstant, Amer. J. Math. **85** 327 (1963).

4 J.P. Serre, *Algèbres de Lie Semisimples Complexes* (Benjamin, New York, 1966); A. Borel, *Linear Algebraic Groups* (Benjamin, New York, 1969).

5 D. Montgomery, *Bombay Colloquium on Differential Analysis* (Oxford U.P., New York, 1964).

6 Séminaire Sophus Lie (1954–1955); P. Cartier, Exposé 19; A.J. Coleman, Can. J. Math., **10**, 349 (1958).

7 L. O'Raifeartaigh, in *Symposia on Theoretical Physics*, no. 2, edited by A. Ramakrishnan (Plenum, New York, 1966).

8 Let $x_0 \in \mathcal{G}$. Then the stratum of x_0 is the set $\{x \in \mathcal{G} \mid \mathcal{G}_x$ is conjugated to $\mathcal{G}_{x_0}$ in $\mathcal{G}\}$.

9 J. Dieudonné, *Foundations of Modern Analysis* (Academic, New York, 1960).

10 V.S. Varadarajan, Amer. J. Math. **90**, 308 (1968).

11 We must take in mind that Sec. 2 deals with $SU(2)$. But $U(2) = U(1) \times SU(2)$ is reductive and its semisimple part is $SU(2)$.

Analytic Treatment of the Coulomb Potential in the Path Integral Formalism by Exact Summation of a Perturbation Expansion

M. J. Goovaerts*† and J. T. Devreese†
Institute for Applied Mathematics, Faculty of Science, University of Antwerp, Middelheimlaan 1, Belgium
(Received 1 March 1971)

A straightforward analytical calculation of the s-like energy spectrum of the hydrogen atom is performed entirely within Feynman's path integral formalism. For this purpose the integral transform $W = \int d^3 \mathbf{r}_\beta K(\mathbf{r}_\beta, 0)$, where $K(\mathbf{r}_\beta, 0)$ is the density matrix of the hydrogen atom written as a path integral, is calculated by means of the exact summation of a "modified" perturbation expansion (W is expanded as a power series in $\sqrt{\beta}$ with $\beta = 1/kT$). Performing this summation is equivalent to solving a problem of moments with infinite moments. For a wide class of potentials the perturbation expansion for W converges faster than the power-series expansion for the exponential function (for the Coulomb potential the convergence rate of both expansions is the same). It is shown how the complete energy spectrum can be obtained by this method. It is also illustrated how the wavefunctions might be obtained by transforming W.

1. INTRODUCTION

In the formulation of quantum mechanics by use of Feynman path integrals[1] exposed, e.g., in Feynman and Hibbs's textbook,[2] two major topics which comprise vital parts of current textbooks on quantum mechanics are not treated: the hydrogen atom and the spin problems. Although formal and numerical progress[3-6] on both problems has been made since 1965, both the spin and the hydrogen atom problems are still unsolved: no direct, analytical solution has been given which is entirely self-consistent within the path integral formalism.

In this paper we will present a direct and self-consistent treatment of the hydrogen atom, using the path integral formalism. The calculations are limited here to the s-like states; but it is shown how the non-s-like states can be treated.

Let us briefly review the literature on path integral calculations concerning the Coulomb potential problem. Storer[3] has calculated the two-particle density matrix for attractive Coulomb forces. This treatment, however, is numerical. Gelman and Spruch[4] have studied scattering theory for central potentials (including the Coulomb potential) in the path integral formalism. They obtained variational upper bounds for the scattering length. The only exact analytical results concerning the hydrogen atom, involving to some extent path integral considerations, are due to Gutzwiller,[5] who applied the WKB approximation in momentum space to the problem. Gutzwiller's paper provides us with a WKB approximation which is manifestly adequate to treat bound states in atoms. Calculating the Green's function $G(\mathbf{p}'', \mathbf{p}', E)$ (i.e., the probability amplitude for a particle with energy E to go from a state with momentum $\mathbf{p}'$ to one with momentum $\mathbf{p}''$) in the limit $\hbar \to 0$, which approximation results in a summation over all classical paths, Gutzwiller obtained poles and residues for the Green's function which exactly represent the energies and wavefunctions of the bound states of the H atom. However Gutzwiller's treatment does not meet our purpose of a direct, exact, analytical study of the hydrogen atom in the path integral formalism. The reasons are the following:

(i) Gutzwiller's treatment is an approximation (WKB). Even if the results which are obtained are exact for the bound-state energies and wavefunctions, this is not proved in the framework of the approximation. It is well known that the WKB approximation gives, e.g., the exact solutions for the linear harmonic oscillator, but again this has not been proved on *a priori* grounds though it is quoted as an accident in standard textbooks on quantum mechanics.[7]

(ii) The author has not examined the scattering states obtained by his approximation. These states contain an important part of the problem.

(iii) The treatment of Gutzwiller is different from a direct path integral calculation because only the limit $\hbar \to 0$ is considered, which leads to a sum over all classical paths instead of a sum over all paths.

(iv) Finally the paths considered by this author are of the Garrod–Feynman[8] type; i.e., these are paths different from the paths in coordinate space originally introduced by Feynman.

It is thus seen that as far as Coulomb potentials are concerned, path integrals have recently played a role in studying scattering theory,[4] distribution functions for electrons and ions in a plasma,[3] and approximation methods for atomic and molecular physics. A direct, exact, analytic solution for the hydrogen atom in the path integral formalism, which we wish to present here, has not been presented up till now.

Given this general expression for W, a mathematical theorem of Grosjean[9] concerning the development of Laplace transforms in the neighborhood of the origin is very helpful to transform the modified perturbation expansion into closed analytical form and to separate the contributions of W belonging to the discrete, or the continuous, part of the spectrum.

For examples like the hydrogen atom, the harmonic oscillator, and the δ-function potential it is possible to perform the summation of the modified perturbation expansion analytically, as we will show here and in Ref. 10. For more general potentials (e.g., the polaron potential) it is possible to determine the energy spectrum by considering the asymptotic behavior of the modified expansion for W for large values of β.[11]

The present paper is divided as follows: In Sec. 2 the modified perturbation expansion for W, starting from the path integral expression, is developed in a straightforward manner resulting in a series of type (1.3) with A_n defined as in (1.4). In Sec. 3 this general technique is applied to the hydrogen atom. The different steps in the general method: introduction of the Fourier transform of the potential, integration over $\mathbf{r}_\beta$, integration over the time variables, are illustrated for the hydrogen atom.

This leads to an expression for W in which the coefficients of $\beta^{n/2}$ are 3n-tuple integrals of a new type. It is shown how these integrals can be performed analytically.

Using Grosjean's theorem the summation of W is then performed. Finally W is written in a closed form in which the contributions from the continuous spectrum and the discrete spectrum are separated.

In what follows we briefly outline the method which we will use here. This method was developed during our calculations on the self-energy and the effective mass of Fröhlich polarons.[11] The polaron potential, however, is closely related to the Coulomb potential, and therefore it is very useful to dispose of an analytic treatment of the hydrogen atom in the framework of path integrals. Especially for the study of the bound polaron in the path integral formalism it is necessary to dispose of a treatment of the hydrogen atom in this formalism. Consider the path integral expression for the one-particle density matrix:

$$\rho(\mathbf{r}_\beta, \mathbf{r}_0) = \int \mathfrak{D}r(t) \exp\left[-\int_0^\beta \left(\frac{m\dot{\mathbf{r}}^2}{2} + V(\mathbf{r}(t))\right) dt\right], \tag{1.1}$$

$\mathbf{r}_\beta$ corresponds to $t = \beta$, $\mathbf{r}_0$ to $t = 0$ (where it is recalled that t is not an actual time variable, but a so-called imaginary time variable), and $\beta = kT$. The notation $\int \mathfrak{D}\mathbf{r}/(t)$ means integration over all paths

It will turn out that the calculations are substantially simplified if, instead of $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$, we calculate the integral transform

$$W = \int \rho(\mathbf{r}_\beta, 0) d^3\mathbf{r}_\beta . \tag{1.2}$$

Of course, some information is lost by treating this expression instead of $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$. States with wavefunctions for which $\psi(\mathbf{r}_\beta) = -\psi(-\mathbf{r}_\beta)$, e.g., not longer, contribute to W. This means that only s states will contribute to the expression W for the hydrogen atom. On the other hand, as we illustrate for the problem of the δ function,[10] the function W occasionally has certain advantages over $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$; the method of continued fractions can be used for the harmonic oscillator to calculate W (via a Laplace transform) starting from the different moments. These moments are finite in the case of W, but the corresponding moments for ρ are divergent. Furthermore it is possible to generalize our treatment to include the complete set of eigenstates of the hydrogen atom, as we will discuss further on.

First the functional $\exp[-\int_0^\beta V(\mathbf{r})]dt$ in (2) will be expanded in a power series in $\int_0^\beta V(\mathbf{r})dt$. It is convenient if V can be expressed as a Fourier transform (although our treatment is more general, as will be shown in the case of the harmonic oscillator). A method is then developed to expand W in powers of $\sqrt{\beta}$. In general, the expansion of W in the neighbourhood of the origin with respect to β will be of the type

$$W = \sum_{n:0}^{\infty} A_n \beta^{n+\sigma-1}, \tag{1.3}$$

with

$$A_n = \alpha_n + \beta^{1-\sigma}\beta_n, \quad 0 < \sigma < 1. \tag{1.4}$$

Because the development is not a power series in β, we will call our expansion for W "a modified perturbation expansion." (The word "perturbation" will not imply that V or $\int_0^\beta V dt$ or β^{-1} is small compared to any characteristic energy.) Each term in the modified perturbation expansion requires the evaluation of a Gaussian path integral, which can be easily performed by using Feynman's original techniques. The integrations over the "time" variables are performed by using the Laplace transformation. At this point one must try to transform the modified perturbation expansion into closed form. One knows that the density matrix can be written as

$$\rho(\mathbf{r}_\beta, \mathbf{r}_0) = \int \exp(-\beta E) d\psi.$$

After integration with respect to $\mathbf{r}_\beta$, we have, under general conditions, for the form of W

$$W = \int \exp(-\beta E) d\psi.$$

In Appendices A–C it is shown that all the information about the energy spectrum and the wavefunctions of the s-like states of the H atom is contained in our expression for W. To obtain the wavefunctions in general, the method has to be generalized slightly.[11]

2. THE MODIFIED PERTURBATION EXPANSION

The density matrix for a particle in a potential $V(\mathbf{r})$ can be written as

$$\rho(\mathbf{r}_\beta, \mathbf{r}_0) = \int \exp\left(-\tfrac{1}{2}\int_0^\beta \dot{\mathbf{r}}^2 dt - \int_0^\beta V(\mathbf{r})dt\right)\mathfrak{D}\mathbf{r}(t), \tag{2.1}$$

where $\mathbf{r}_\beta$ corresponds to $t = \beta$ and $\mathbf{r}_0$ to $t = 0$. Expansion of the exponential with respect to $\int_0^\beta V(\mathbf{r})dt$ gives

$$\rho(\mathbf{r}_\beta, \mathbf{r}_0) = \int \exp\left(-\tfrac{1}{2}\int_0^\beta \dot{\mathbf{r}}^2 dt\right) \sum_{n:0}^{\infty} \frac{(-1)^n}{n!}\left(\int_0^\beta V(\mathbf{r})dt\right)^n \times \mathfrak{D}\mathbf{r}(t). \tag{2.2}$$

If one writes the nth power of the integral as an n-dimensional integral and introduces the Fourier transform

$$V(\mathbf{r}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \exp(i\mathbf{k}\cdot\mathbf{r}) f(\mathbf{k}), \tag{2.3}$$

one obtains the following development for $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$:

$$\rho(\mathbf{r}_\beta, \mathbf{r}_0) = \sum_{n:0}^{\infty} \frac{(-1)^n}{n!} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \int \frac{d^3\mathbf{k}_1}{(2\pi)^3} f(\mathbf{k}_1) \cdots \times \int \frac{d^3\mathbf{k}_n}{(2\pi)^3} f(\mathbf{k}_n) \int \exp\left(-\tfrac{1}{2}\int_0^\beta \dot{\mathbf{r}}^2 dt + \times\, i \sum_{j:1}^{n} \mathbf{r}(t_j)\mathbf{k}_j\right)\mathfrak{D}\mathbf{r}(t). \tag{2.4}$$

Defining an extension of the function $\mathbf{f}(t)$, introduced by Feynman,[12] as

$$\mathbf{f}_n(t) = i \sum_{j:1}^{n} \delta(t - t_j)\mathbf{k}_j, \tag{2.5}$$

one still has to calculate the path integrals

$$P_n = \int \exp\left(-\tfrac{1}{2}\int_0^\beta \dot{\mathbf{r}}^2 dt + \int_0^\beta \mathbf{f}_n(t)\cdot\mathbf{r}(t)dt\right) \mathfrak{D}\mathbf{r}(t), \quad n: 0, 1, 2, \cdots. \tag{2.6}$$

These path integrals are Gaussian and can easily be calculated. Some of the details are given in Appendix A. The result is

$$P_n = (2\pi\beta)^{-3/2} \exp(S_{cl,n}), \tag{2.7}$$

with

$$\begin{aligned} S_{cl,n} = &-\frac{1}{2\beta}(\mathbf{r}_\beta - \mathbf{r}_0)^2 + i\sum_{j:1}^{n}\mathbf{k}_j \frac{t_j}{\beta}(\mathbf{r}_\beta - \mathbf{r}_0) \\ &+ i\mathbf{r}_0 \sum_{j:1}^{n}\mathbf{k}_j - \tfrac{1}{2}(\mathbf{k}_1^2 T'_{1,1} + \cdots + \mathbf{k}_n^2 T'_{n,n} \\ &+ 2\mathbf{k}_1\mathbf{k}_2 T'_{1,2} + \cdots + 2\mathbf{k}_{n-1}\mathbf{k}_n T'_{n-1,n}). \end{aligned} \tag{2.8}$$

$T'_{i,j}$ is defined as

$$T'_{i,j} = -(t_i t_j/\beta) + \tfrac{1}{2}(t_i + t_j) - \tfrac{1}{2}|t_i - t_j| \tag{2.9}$$

As far as the integrations over the variables t are concerned, the difficulties arise from the products $t_i t_j$, because of which the argument of the exponential function is not linear in the "time" variables. In the expression for W the cumbersome product terms $t_i t_j/\beta$ no longer appear. However, states for which the probability amplitude at the origin or the average of the probability amplitude over all space is zero are no longer contained in the expression for W. (For the hydrogen atom this would mean that only s states appear in W, but we will show the treatment can be generalized.)

As was discussed in the Introduction, we introduce the expression

$$W = \int d^3\mathbf{r}_\beta \rho(\mathbf{r}_\beta, \mathbf{r}_0).$$

Integration over $\mathbf{r}_\beta$ is elementary. If the series expansion for ρ (2.4) is now introduced, one has to calculate

$$W'_n = \int d^3\mathbf{r}_\beta \int \exp\left[\left(\tfrac{1}{2}\int_0^\beta \dot{\mathbf{r}}^2 dt + i\sum_{j:1}^{n}\mathbf{r}(t_j)\mathbf{k}_j\right)\right]\mathfrak{D}\mathbf{r}(t),$$

and finds

$$W'_n = \exp\{-\tfrac{1}{2}[K_n]^T[T_n][K_n]\},$$

where $[K_n]$ is a column vector defined as

$$[K_n] = \begin{bmatrix}\mathbf{k}_1\\ \vdots \\ \mathbf{k}_n\end{bmatrix},$$

and $[T_n]$ is a square matrix of order n, with elements defined as

$$T_{i,j} = \tfrac{1}{2}(t_i + t_j) - \tfrac{1}{2}|t_i - t_j|. \tag{2.10}$$

W can now be written as

$$W = \sum_{n:0}^{\infty} \frac{(-1)^n}{n!}\frac{\sqrt{2}^{3n}}{(2\pi)^{3n}} \int d^3\mathbf{k}_1 f(\mathbf{k}_1\sqrt{2}) \cdots \int d^3\mathbf{k}_n \times f(\mathbf{k}_n\sqrt{2}) \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \exp\{-[K_n]^T[T_n][K_n]\}. \tag{2.11}$$

First the integration over the time variables is treated. Consider

$$H_n(\beta) = \int_0^\infty dt_1 \cdots \int_0^\beta dt_n \exp\{-[K_n]^T[T_n][K_n]\}. \tag{2.12}$$

From the symmetry properties of the integrandum it follows that

$$H_n(\beta) = n!\int_0^\beta dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \times \exp\{-[K_n]^T[T_n][K_n]\}.$$

For the Laplace transform of $H_n(\beta)$ one finds

$$\mathfrak{L}\left[\frac{H_n(s)}{n!}, s\right] = \frac{1}{s[s + k_1^2][s + (\mathbf{k}_1 + \mathbf{k}_2)^2]\cdots[s + (\mathbf{k}_1 + \mathbf{k}_2 + \cdots + \mathbf{k}_n)^2]},$$

so that the modified perturbation expansion becomes

$$W = \mathfrak{L}^{-1}\sum_{n:0}^{\infty}(-1)^n \frac{\sqrt{2}^{3n}}{(2\pi)^{3n}} \int d^3\mathbf{k}_1 f(\sqrt{2}\,\mathbf{k}_1)\cdots \int d^3\mathbf{k}_n f(\sqrt{2}\mathbf{k}_n)\frac{1}{s(s + \mathbf{k}_1^2)\cdots s + (\mathbf{k}_1 + \mathbf{k}_2 + \cdots + \mathbf{k}_n)^2}.$$

The inverse Laplace transform can easily be calculated in general. In the special case of a potential which is a homogeneous function of order h, the expansion becomes

$$W = \sum_{n:0}^{\infty}(-1)^n \frac{1}{(2\pi)^{3n}}\frac{1}{\sqrt{2}^{hn}}\frac{\beta^{(h+2)n/2}}{\Gamma[(h+2)\frac{1}{2}n + 1]}\int \frac{d^3\mathbf{k}_1 f(\mathbf{k}_1)}{(1 + \mathbf{k}_1^2)}\int \frac{d^3\mathbf{k}_2 f(\mathbf{k}_2)}{1 + (\mathbf{k}_1 + \mathbf{k}_2)^2}\cdots\frac{d^3\mathbf{k}_n f(\mathbf{k}_n)}{[s + (\mathbf{k}_1 + \mathbf{k}_2 + \cdots + \mathbf{k}_n)^2]}. \tag{2.13}$$

[This expression is of the same type as the Liouville solution of the integral equation for the Green's function $G(\mathbf{p}, \mathbf{p}', \omega)$ in the momentum representation. However W is not directly expressible in terms of $G(\mathbf{p}, \mathbf{p}', \omega)$. Also this expression has been derived from the path integral(2.1) in a straightforward manner.]

If $V(\mathbf{o}) \neq \infty$ this series is absolutely convergent for every finite β. The convergence is stronger than that of the power series development of the exponential function. If $V(\mathbf{o}) = \infty$, the expansion may be convergent (examples are the δ function potential and the hydrogen atom). Further discussions of more complicated expansions of type (2.13) are given in Ref. 10.

In view of Grosjean's theorem,[9] expression (2.13) has the general form one would expect a priori for both the density matrix and W. Expansion (2.13) contains information about both the discrete and continuous spectrum as shown in Ref. 13.

3. APPLICATION OF THE MODIFIED PERTURBATION EXPANSION TO THE CASE OF THE COULOMB-POTENTIAL

A. Perturbation Expansion of the Density Matrix with Respect to the Coulomb Potential

The potential is given by

$$V(\mathbf{r}) = -e^2/|\mathbf{r}|. \tag{3.1}$$

Writing the Coulomb potential as a Fourier transform results in

$$V(\mathbf{r}) = \frac{-e^2}{|\mathbf{r}|} = -e^2 \int \frac{d^3\mathbf{k}}{2\pi^2\mathbf{k}^2} \exp(i\mathbf{k}\mathbf{r}). \tag{3.2}$$

The Coulomb potential is an example of the case $V(\mathbf{o}) = \infty$.

Expansion (2.4) becomes

$$\rho(\mathbf{r}_\beta, \mathbf{r}_0) = \sum_{n:0}^{\infty} \frac{(-1)^n}{n!} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \int \frac{d^3\mathbf{k}_1}{2\pi^2\mathbf{k}_1^2} \cdots \times \int \frac{d^3\mathbf{k}_n}{2\pi^2\mathbf{k}_n^2} \frac{e^{S_{cl,n}}}{(2\pi\beta)^{3/2}}. \tag{3.3}$$

The zeroth order ($n = 0$) is equal to the kernel of the free particle with boundary conditions

$$\begin{aligned} t &= 0, \quad \mathbf{r}(0) = \mathbf{r}_0, \\ t &= \beta, \quad \mathbf{r}(\beta) = \mathbf{r}_\beta, \end{aligned} \tag{3.4}$$

B. The Modified Perturbation Expansion of the Integral Transform W of the Density Matrix

With (2.11) the integral transform W takes the form

$$W = \sum_{n:0}^{\infty} \frac{e^{2n}}{n!} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \int \frac{d^3\mathbf{k}_1}{2\pi^2\mathbf{k}_1^2} \cdots \int \frac{d^3\mathbf{k}_n}{2\pi^2\mathbf{k}_n^2} \times \exp\{-\tfrac{1}{2}[K_n]^T[T_n][K_n]\}, \tag{3.5}$$

which will be written as

$$W = \sum_{n:0}^{\infty} e^{2n} C_n(\beta), \tag{3.6}$$

with the notation

$$C_n(\beta) = \frac{1}{n!} \frac{1}{2^n} \frac{\sqrt{2}^n}{\pi^{2n}} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \int \frac{d^3\mathbf{k}_1}{\mathbf{k}_1^2} \cdots \times \int \frac{d^3\mathbf{k}_n}{\mathbf{k}_n^2} \exp\{-[K_n]^T[T_n][K_n]\}. \tag{3.7}$$

C. Integration over the k Vectors

In the expression for $C_n(\beta)$ the Cartesian coordinates of the different vectors cannot be separated. Therefore, the following integral transformation is used;

$$\frac{1}{\mathbf{k}_1^2} = \int_0^\infty e^{-\epsilon_1 k_1^2} d\epsilon_1. \tag{3.8}$$

Written explicitly $C_n(\beta)$ becomes

$$C_n(\beta) = \frac{1}{n!} \frac{1}{2^n} \frac{\sqrt{2}^n}{\pi^{2n}} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \int_0^\infty d\epsilon_1 \cdots \times \int_0^\infty d\epsilon_n \int d^3\mathbf{k}_1 \cdots \int d^3\mathbf{k}_n \exp\{-[\mathbf{k}_1^2(\epsilon_1 + T_{1,1}) + \cdots + \mathbf{k}_n^2(\epsilon_n + T_{n,n}) + 2\mathbf{k}_1\mathbf{k}_2 T_{1,2} + \cdots + 2\mathbf{k}_{n-1}\mathbf{k}_n T_{n-1,n}]\}.$$

For this expression the Cartesian coordinates can be separated as follows:

$$C_n(\beta) = \frac{1}{n!} \frac{1}{2^n} \frac{\sqrt{2}^n}{\pi^{2n}} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \int_0^\infty d\epsilon_1 \cdots \int_0^\infty d\epsilon_n \times \left(\int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n \exp\{-[x_1^2(\epsilon_1 + T_{1,1}) + \cdots + 2x_{n-1} x_n T_{n-1,n}]\}\right)^3.$$

The integral

$$K_n = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n \exp[-(x_1^2 B_{1,1} + \cdots + x_n^2 B_{n,n} + 2x_1x_2B_{1,2} + \cdots + 2x_{n-1}x_nB_{n-1,n})] \tag{3.9}$$

is a standard n tuple integral and equals

$$K_n = \frac{\pi^{n/2}}{\{\det[B_n]\}^{1/2}}. \tag{3.10}$$

$[B_n]$ is the matrix with elements $B_{i,j} = \epsilon_i \delta_{i,j} + T_{i,j}$. The only condition is $\det[B_n] > 0$, which in the present case is sufficient to show that $[B_n]$ is positive definite. It is shown now that $\det[B_n]$ is positive except for some values of the time variables where $\det[B_n] = 0$. In the case $n = 2$ one has

$$[B_2] = \begin{bmatrix} \epsilon_1 + t_1 & \tfrac{1}{2}(t_1 + t_2) - \tfrac{1}{2}|t_1 - t_2| \\ \tfrac{1}{2}(t_1 + t_2) - \tfrac{1}{2}|t_1 - t_2| & \epsilon_2 + t_2 \end{bmatrix}.$$

If $t_1 \geqslant t_2$, it follows that

$$\det[B_2] = (\epsilon_1 + t_1)(\epsilon_2 + t_2) - t_2^2 \geqslant t_1 t_2 - t_2^2 \geqslant 0.$$

The exceptional points for which $\det[B_2] = 0$ are

$$\epsilon_1 = \epsilon_2 = 0, \quad t_2 = 0, \text{and } \epsilon_1 = \epsilon_2 = 0, \quad t_1 = t_2.$$

The fact that $\det[B_n] = 0$ in a finite number of points does not affect the value of K_n. Analogous results hold in the complementary case $t_2 \geqslant t_1$.

By means of complete induction it is easily shown that $[B_n]$ is positive definite for all n except for a finite number of points, for each n, where $\det[B_n] = 0$. The problem of finding analytical results for the coefficients in the modified perturbation expansion is reduced to calculating

$$C_n(\beta) = \frac{1}{n!}\frac{1}{\sqrt{2^n}}\frac{\pi^{3n/2}}{\pi^{2n}}\int_0^\beta dt_1\cdots\int_0^\beta dt_n\int_0^\infty d\epsilon_1\cdots \times\int_0^\infty d\epsilon_n\,\frac{1}{\det[B_n]^3}. \tag{3.11}$$

D. Transformation and Evaluation of the Integrations over the Time Variables $t_1\cdots t_n$

As in the general case, some symmetry properties of the integrandum can be used leading to the equality

$$C_n(\beta) = \frac{1}{\sqrt{2^n}}\frac{\sqrt{\pi}^{3n/2}}{\pi^{2n}}\int_0^\beta dt_1\int_0^{t_1} dt_2\cdots\int_0^{t_{n-1}} dt_n\int_0^\infty d\epsilon_1\cdots\int_0^\infty d\epsilon_n\,\frac{1}{\det[B_n]^3}, \tag{3.12}$$

with

$$\det[B_n] = \begin{vmatrix} \epsilon_1+t_1 & t_2 & t_3 & \cdots & \cdots & t_n \\ t_2 & \epsilon_2+t_2 & \cdots & & & \\ \vdots & \vdots & & & & \vdots \\ \vdots & & & & & \vdots \\ t_n & & & \cdots & t_n & \epsilon_n+t_n \end{vmatrix}.$$

Using basic properties of determinants it follows that

$$\det[B_n] = \begin{vmatrix} t_1-t_2+\epsilon_1+\epsilon_2 & -\epsilon_2 & 0 & 0 \\ -\epsilon_2 & t_2-t_3+\epsilon_2+\epsilon_3 & & \\ 0 & & & \\ \vdots & & & \vdots \\ & & \epsilon_n+\epsilon_{n-1}+t_{n-1}t_n & -\epsilon_n \\ 0 & \cdots & -\epsilon_n & \epsilon_n+t_n \end{vmatrix}. \tag{3.13}$$

From (3.9) and (3.10) combined with (3.13), the following integral representation for $\{\det[B_n]\}^{-3/2}$ is derived:

$$(\det B_n)^{-3/2} = \pi^{-3n/2}\int d^3\mathbf{k}_1\cdots\int d^3\mathbf{k}_n\,\exp\{-[\mathbf{k}_1^2(t_1-t_2+\epsilon_1+\epsilon_2)+\cdots+\mathbf{k}_{n-1}^2(t_{n-1}-t_n+\epsilon_{n-1}+\epsilon_n) + \mathbf{k}_n^2(\epsilon_n+t_n) - 2\epsilon_2\mathbf{k}_1\mathbf{k}_2\cdots 2\epsilon_n\mathbf{k}_{n-1}\mathbf{k}_n\}. \tag{3.14}$$

To evaluate the integrals over the time variables a Laplace transformation with respect to β is used. Putting

$$Q_n(s) = \int_0^\infty \exp(-\beta s)d\beta\int_0^\beta dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n\,\exp[-\mathbf{k}_1^2(t_1-t_2)-\mathbf{k}_2^2(t_2-t_3)\cdots\mathbf{k}_{n-1}^2(t_{n-1}-t_n)-\mathbf{k}_n^2t_n]$$

one finds, by means of partial integration,

$$Q_n(s) = \frac{1}{s(s+\mathbf{k}_1^2)\cdots(s+k_n^2)},$$

so that the Laplace transform of $C_n(\beta)$ becomes

$$\mathfrak{L}[C_n(\beta),s] = \frac{1}{\sqrt{2^n}}\frac{1}{\pi^{2n}}\int_0^\infty d\epsilon_1\cdots\int_0^\infty d\epsilon_n\int\frac{d^3\mathbf{k}_1}{(s+\mathbf{k}_1^2)}\cdots\int\frac{d^3\mathbf{k}_n}{(s+\mathbf{k}_n^2)}\frac{1}{s}\exp\{-[k_1^2(\epsilon_1+\epsilon_2)\cdots+\mathbf{k}_{n-1}^2(\epsilon_{n-1}+\epsilon_n) + \mathbf{k}_n^2\epsilon_n - 2\epsilon_2\mathbf{k}_1\mathbf{k}_2-\cdots-2\epsilon_n\mathbf{k}_{n-1}\mathbf{k}_n]\}.$$

The integrations over the variables $\epsilon_1\cdots\epsilon_n$ are elementary now:

$$\mathfrak{L}[C_n(\beta),s] = \frac{1}{s}\frac{1}{\sqrt{2^n}}\frac{1}{\pi^{2n}}\int\frac{d^3\mathbf{k}_1}{(s+\mathbf{k}_1^2)}\cdots\int\frac{d^3\mathbf{k}_n}{(s+\mathbf{k}_n^2)}\frac{1}{\mathbf{k}_1^2}\frac{1}{(\mathbf{k}_2-\mathbf{k}_1)^2}\cdots\frac{1}{(\mathbf{k}_n-\mathbf{k}_{n-1})^2}. \tag{3.15}$$

The evaluation of the inverse Laplace transform in both members of (3.15) is readilly performed. One gets the expression

$$C_n(\beta) = \frac{\beta^{n/2}}{\Gamma(\frac{1}{2}n+1)} \frac{1}{\sqrt{2^n}} \frac{1}{\pi^{2n}} \int \frac{d^3\mathbf{k}_1}{(1+\mathbf{k}_1^2)} \cdots \int \frac{d^3\mathbf{k}_n}{(1+\mathbf{k}_n^2)} \frac{1}{\mathbf{k}_1^2} \frac{1}{(\mathbf{k}_2-\mathbf{k}_1)^2} \cdots \frac{1}{(\mathbf{k}_n-\mathbf{k}_{n-1})^2}. \qquad (3.16)$$

[Again the similarity with the iterated solution of the corresponding integral equation for the Green's function $G(\mathbf{p}_2, \mathbf{p}_1, \omega)$ should be remarked[14]:

$$G(\mathbf{p}_2, \mathbf{p}_1, \omega) = \frac{\delta^3(\mathbf{p}_2 - \mathbf{p}_1)}{\mathbf{p}^2 - \mathbf{p}_2^2} - \frac{2p\upsilon}{p^2 - p_2^2} \int \frac{d^3\mathbf{p}_3}{(2\pi)^3} \frac{4\pi}{|\mathbf{p}_2 - \mathbf{p}_3|^2} G(\mathbf{p}_3, \mathbf{p}_2, \omega), p = (2m\omega/h)^{1/2}, \upsilon = Ze^2m/4\pi p h^2.]$$

Introducting spherical coordinates, the integrations over the angles can be carried out immediately to give

$$C_n(\beta) = \frac{\beta^{n/2}}{\Gamma^{1/2}(n+1)} \frac{1}{\sqrt{2^n}} \frac{1}{\pi^{2n}} (2\pi)^{n/2} \int_0^\infty \frac{k_1^2 dk_1}{(1+k_1^2)} \cdots \int_0^\infty \frac{k_n^2 dk_n}{(1+k_n^2)} \frac{1}{k_1^2} \frac{1}{k_1k_2} \frac{1}{k_2k_3} \cdots \frac{1}{k_{n-1}k_n}$$
$$\times \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| \cdots \ln\left|\frac{k_{n-1}+k_n}{k_{n-1}-k_n}\right|. \qquad (3.17)$$

E. Calculation of the n-tuple Integral $Z_n (n \geqslant 2)$

The n tuple integral in (3.17) is called Z_n:

$$Z_n = \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \int_0^\infty \frac{dk_2}{(1+k_2^2)} \cdots \int_0^\infty \frac{dk_{n-1}}{(1+k_{n-1}^2)}$$
$$\times \int_0^\infty \frac{k_n dk_n}{(1+k_n^2)} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| \cdots \ln\left|\frac{k_{n-1}+k_n}{k_{n-1}-k_n}\right|. \qquad (3.18)$$

We treat this integral for $n=2, n=3$, and still the general case.

1. Special Case $n=2$

For $n=2$ one has to calculate

$$Z_2 = \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \int_0^\infty \frac{k_2 dk_2}{(1+k_2^2)} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right|.$$

Using complex integration one has the result

$$\int_0^\infty \frac{k_2 dk_2}{1+k_2^2} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| = \pi \arctan k_1, \qquad (3.19)$$

so that Z_2 is reduced to

$$Z_2 = \pi \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \arctan k_1.$$

Or, using an integral representation for $\arctan k_1$, we obtain

$$Z_2 = \pi \int_0^1 dy_1 \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \frac{k_1}{(1+y_1^2k_1^2)}.$$

Integration over the variable k_1 is elementary now and results in

$$Z_2 = \pi \int_0^1 dy_1 \frac{\pi}{2} \frac{1}{1+y_1}. \qquad (3.20)$$

Written in analytic form, this integral takes the form

$$Z_2 = \frac{\pi^2}{2} \ln 2 = \frac{\pi^2}{2}\left(1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \cdots + \frac{(-1)^{n+1}}{n} \cdots\right).$$

2. Special Case $n=3$

In this case one has

$$Z_3 = \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \int_0^\infty \frac{dk_2}{(1+k_2^2)} \int_0^\infty \frac{k_3 dk_3}{(1+k_3^2)}$$
$$\times \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| \ln\left|\frac{k_2+k_3}{k_2-k_3}\right|. \qquad (3.21)$$

Integration over k_3 is equivalent to the integration performed in (3.19). For Z_3 one obtains the expression

$$Z_3 = \pi \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \int_0^\infty \frac{dk_2}{(1+k_2^2)} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right|$$
$$\arctan k_2. \qquad (3.22)$$

Again using an integral representation of $\arctan k$, (3.22) becomes

$$Z_3 = \pi \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \int_0^1 dy_2 \int_0^\infty \frac{k_2 dk_2}{(1+k_2^2)}$$
$$\times \frac{1}{(1+y_2^2k_2^2)} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right|. \qquad (3.23)$$

By writing $1/(1+k_2^2)(1+y_2^2k_2^2)$ as a sum of two partial fractions, one arrives, after a few elementary calculations, at

$$\int_0^\infty \frac{k_2 kd_2}{1+k_2^2} \frac{1}{1+y_2^2k_2^2} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right|$$
$$= \frac{\pi}{1-y_2^2} [\arctan k_1 - \arctan k_1 y_2],$$

which can be written as

$$\frac{\pi}{1-y_2^2} \int_{y_2}^1 dy_1 \int_0^\infty \frac{k_1 dk_1}{1+y_1^2k_1^2},$$

so that Z_3 becomes

$$Z_3 = \pi^2 \int_0^1 \frac{dy_2}{1-y_2^2} \int_{y_2}^1 dy_1 \int_0^\infty \frac{dk_1}{(1+k_1^2)(1+y_1^2k_1^2)}$$

or

$$Z_3 = \frac{\pi^3}{2} \int_0^1 \frac{dy_2}{1-y_2^2} \int_{y_2}^1 \frac{dy_1}{1+y_1} .$$

In analytic form this is

$$Z_3 = \frac{\pi^3}{4} \left(1 - \frac{1}{2^2} + \frac{1}{3^2} - - \frac{1}{4^2} + \cdots\right).$$

3. General Case of Z_n

Carrying out the integration over k_n produces

$$Z_n = \pi \int_0^\infty \frac{dk_1}{k(1+k_1^2)} \int_0^\infty \frac{dk_2}{(1+k_2^2)} \cdots \int_0^\infty \frac{dk_{n-1}}{(1+k_{n-1}^2)}$$
$$\times \arctan k_{n-1} \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| \cdots \ln\left|\frac{k_{n-2}+k_{n-1}}{k_{n-2}-k_{n-1}}\right|.$$

Using the same integral representation as in the cases $n = 2$ and 3 and carrying out the integration over k_{n-1}, one finds

$$Z_n = \pi^2 \int_0^1 \frac{dy_{n-1}}{1-y_{n-1}^2} \int_{y_{n-1}}^1 dy_{n-2} \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)}$$
$$\times \int_0^\infty \frac{dk_2}{(1+k_2^2)} \cdots \int_0^\infty \frac{dk_{n-2}k_{n-2}}{(1+k_{n-2}^2)(1+y_{n-2}^2k_{n-2}^2)}$$
$$\times \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| \cdots \ln\left|\frac{k_{n-3}+k_{n-2}}{k_{n-3}-k_{n-2}}\right|.$$

Performing systematically the integrations over $k_{n-2} \cdots k_3$, following the same rules, results in

$$Z_n = \pi^{n-2} \int_0^1 \frac{dy_{n-1}}{1-y_{n-1}^2} \int_{y_{n-1}}^1 \frac{dy_{n-2}}{1-y_{n-2}^2} \cdots \int_{y_4}^1 \frac{dy_3}{1-y_3^2}$$
$$\times \int_{y_3}^1 dy_2 \int_0^\infty \frac{dk_1}{k_1(1+k_1^2)} \int_0^\infty \frac{k_2 dk_2}{(1+k_2^2)(1+y_2^2k_2^2)}$$
$$\times \ln\left|\frac{k_1+k_2}{k_1-k_2}\right| .$$

The remaining integrations over k_1 and k_2 are already described in the case $n = 3$. As a final result for the n tuple integral Z_n we obtain

$$Z_n = \frac{\pi^n}{2} \int_0^1 \frac{dy_{n-1}}{1-y_{n-1}^2} \int_{y_{n-1}}^1 \frac{dy_{n-2}}{1-y_{n-2}^2} \cdots \int_{y_3}^1 \frac{dy_2}{1-y_2^2} \int_{y_2}^1 \frac{dy_1}{1+y_1} . \tag{3.24}$$

With the substitution

$$Y_i = (1 - t_i/1 + t_i)$$

Z_n becomes, successively,

$$Z_n = \frac{\pi^n}{2} \frac{1}{2^{n-2}} \int_0^1 \frac{dt_{n-1}}{t_{n-1}} \int_0^{t_{n-1}} \frac{dt_{n-2}}{t_{n-2}} \cdots \int_0^{t_3} \frac{dt_2}{t_2}$$
$$\times \int_0^{t_2} \frac{dt_1}{1+t_1}$$

and

$$Z_n = \frac{\pi^n}{2} \frac{1}{2^{n-2}} \left(1 - \frac{1}{2^{n-1}} + \frac{1}{3^{n-1}} - \frac{1}{4^{n-1}} + \cdots\right). \tag{3.25}$$

F. Summation of the Modified Perturbation Expansion

Summarizing, one has until now

$$C_0(\beta) = 1,$$

$$C_1(\beta) = \frac{\sqrt{2}\sqrt{\beta}}{\Gamma(\frac{1}{2}+1)} = \frac{2\sqrt{2}\sqrt{\beta}}{\Gamma(\frac{1}{2}+1)} [1 - 1 + 1 - 1 + \cdots],$$

where $(1 - 1 + 1 - 1 \cdots)$ is defined as $\frac{1}{2}$ by Cesaro's summation process, and in general

$$C_n(\beta) = \frac{\beta^{n/2} 2^2}{\Gamma(\frac{1}{2}n+1)\sqrt{2^n}} \left(1 - \frac{1}{2^{n-1}} + \frac{1}{3^{n-1}} + \cdots\right).$$

We have therefore obtained the following expression for the perturbation expansion of W:

$$W = 1 + \sum_{n:1}^\infty \frac{e^{2n}\beta^{n/2}2^2}{\Gamma(\frac{1}{2}n+1)\sqrt{2^n}} \left(1 - \frac{1}{2^{n-1}} + \frac{1}{3^{n-1}} + \cdots\right) \tag{3.26}$$

The free-particle contribution is contained in the term $1(\beta = 0)$. As was already explained in the general case, expansion (3.26) consists of two parts, a power series in β and a power series in $\sqrt{\beta}$; the former is connected with both bound states and scattering states, the latter only with the scattering states.

Taking into account the convergence of both series, one can write

$$W = 1 + \sum_{r:1}^\infty \frac{e^{4r}\beta^{2r}2^2}{\Gamma(r+1)} \frac{1}{2^r} \left(1 - \frac{1}{2^{2r-1}} + \frac{1}{3^{2r-1}} - \frac{1}{4^{2r-1}} + \cdots\right) + \sum_{r:0}^\infty \frac{e^{2(r+1/2)}\beta^{r+1/2}2^2}{\Gamma(r+1+\frac{1}{2})2^{r+1/2}}$$
$$\times \left(1 - \frac{1}{2^{2r+1-1}} + \frac{1}{3^{2r+1-1}} \cdots\right). \tag{3.27}$$

W has the structure of the development of a Laplace transform in the neighborhood of the origin. Calculating the moments of the Laplace transform of W one finds divergencies, so that the theorems of Stieltjes cannot be applied. Performing the summation of this series, however, as will be done here, is equivalent to solving a "problem of moments" with divergent moments. Grosjean[9] has proved the following theorem concerning the development of Laplace transforms in the neighborhood of the origin:

If

$$\frac{a_1}{t^\sigma} + \frac{a_2}{t^{1+\sigma}} + \frac{a_3}{t^{2+\sigma}} + \cdots, \qquad 0 < \sigma < 1,$$
$$a_n: \text{real} \tag{3.28}$$

is either a convergent, or a asymptotic expansion of a function $f(t)$, valid for large t values, then

$$L = \int_0^\infty f(t) e^{-xt} dt \tag{3.29}$$

can be written as

$$L = \sum_{n:0}^{\infty} A_n x^{n+\sigma-1}, \quad x > 0, \tag{3.30}$$

with A_n defined as

$$A_n = (-1)^n \left[\frac{\pi}{\sin \pi\sigma} \frac{a_{n+1}}{\Gamma(n+\sigma)} + \frac{x^{1-\sigma}}{n!} \int_0^\infty \left(t^n f(t) - a_1 t^{n-\sigma} - \cdots - \frac{a_{n+1}}{t^\sigma} \right) dt \right]. \tag{3.31}$$

To perform the summation of the series (3.27) the following problem has to be solved: given A_n, calculate a_n to determine $f(t)$.

In order to apply directly Grosjean's theorem we consider

$$W' = \frac{\partial W}{\partial B} = 2e^4[E + O],$$

with E and O defined as

$$E = \sum_{r:1}^{\infty} \left(\frac{e^4\beta}{2} \right)^{r-1} \frac{1}{\Gamma(r)} \left(1 - \frac{1}{2^{2r-1}} + \frac{1}{3^{2r-1}} \right) \cdots,$$

$$O = \sum_{r:0}^{\infty} \left(\frac{e^4\beta}{2} \right)^{r-1/2} \frac{1}{\Gamma(r+\frac{1}{2})} \left(1 - \frac{1}{2^{2r}} + \frac{1}{3^{2r}} \cdots \right).$$

The series in the general theorem corresponding with O can be written as

$$\overline{O}(x) = \sum_{r:0}^{\infty} x^{r-1/2} (-1)^r \frac{a_{r+1}\pi}{\Gamma(r+\frac{1}{2})}.$$

Putting $x = \frac{1}{2} e^4\beta$, we obtain

$$a_{r+1} = \frac{(-1)^r}{\pi} \left(1 - \frac{1}{2^{2r}} + \frac{1}{3^{2r}} \cdots \right)$$

or

$$a_{r+1} = \frac{(-1)^r}{\pi} \frac{2^{2r-1} - 1}{2^{2r-1}} \zeta(2r), \quad r > 0.$$

With the aid of an integral representation for the ζ function of Riemann, a_{r+1} takes the form

$$a_{r+1} = \frac{(-1)^r}{\pi} \frac{1}{\Gamma(2r+1)} \int_0^\infty \frac{t^{2r} e^t dt}{(e^t+1)^2}, \quad r > 0.$$

It is easy to verify that this expression gives also the correct result in the case $r = 0$.

The behaviour of the series corresponding to (3.28) for large k is

$$\overline{f}(k) \cong \frac{1}{\pi} \sum_{r:0}^{\infty} \frac{1}{k^{r+1/2}} \frac{(-1)^r}{\Gamma(2r+1)} \int_0^\infty \frac{t^{2r} e^t dt}{(e^t+1)^2}.$$

For $k > 1$ this series defines a holomorphic function:

$$\overline{f}(k) \cong \frac{1}{\pi\sqrt{k}} \int_0^\infty \frac{e^t dt}{(e^t+1)^2} \cos \frac{t}{\sqrt{k}}, \quad k > 1.$$

The integral defines a holomorphic function for every $k > 0$. With this integral representation, $f(k)$ is defined:

$$f(k) = \frac{1}{\pi} \frac{1}{k} \int_0^\infty \frac{e^t dt}{(e^t+1)^2} \cos \frac{t}{\sqrt{k}}, \quad k > 0.$$

Using this definition for $f(k)$ we consider a function $R(\beta)$ defined by

$$R(\beta) = \frac{1}{\pi} \int_0^\infty \exp\left(-\frac{e^4\beta k}{2} \right) \frac{dk}{\sqrt{k}} \int_0^\infty \frac{e^t dt}{(e^t+1)^2} \times \cos \frac{t}{\sqrt{k}}. \tag{3.32}$$

From Grosjean's theorem it is immediately clear that the series expansion of (3.32) gives, for the part corresponding with $\overline{O}$, the series O.

For the series expansion in β one has

$$\overline{E}(x) = \sum_{r:0}^{\infty} \frac{x^r(-1)^r}{r!} \int_0^\infty \left(k^r f(k) - a_1 k^{r-1/2} - \cdots - \frac{a_{r+1}}{\sqrt{k}} \right) dk, \tag{3.33}$$

with x equal to $\frac{1}{2} e^4\beta$:

$$b_r = \frac{(-1)^r}{r!} \int_0^\infty \left(k^r f(k) - \sum_{j:0}^{r} a_{j+1} k^{r-j-1/2} \right) dk.$$

Using the definition formulas for a_{r+1} and $f(k)$ one finds, after carrying out some elementary substitutions,

$$b_r = \frac{(-1)^r}{r!} \frac{2}{\pi} \int_0^\infty \frac{d\tau}{\tau^{2r+2}} \left(\cos\tau - \sum_{j:0}^{r} \frac{(-1)^j \tau^{2j}}{(2j)!} \right) \times \int_0^\infty \frac{k^{2n+1} e^k}{(e^k+1)^2} dk. \tag{3.34}$$

After sequential partial integrations, one sees that

$$\int_0^\infty \frac{d\tau}{\tau^{2r+2}} \left(\cos\tau - \sum_{j:0}^{r} \frac{(-1)^j \tau^{2j}}{(2j)!} \right) = \frac{(-1)^{r+1}}{\Gamma(2r+2)} \frac{\pi}{2},$$

while one partial integration leads to

$$\int_0^\infty \frac{k^{2r+1} e^k dk}{(e^k+1)^2} = \Gamma(2r+2) \left(1 - \frac{1}{2^{2r+1}} + \frac{1}{3^{2r+1}} - \frac{1}{4^{2r+1}} \cdots \right),$$

so that

$$b_r = \frac{(-1)^r}{r!} \frac{2}{\pi} \frac{(-1)^{r+1}}{\Gamma(2r+2)} \frac{\pi}{2} \Gamma(2r+2) \times \left(1 - \frac{1}{2^{2r+1}} + \frac{1}{3^{2r+1}} \cdots \right) \tag{3.35}$$

or

$$b_r = -\frac{1}{r!} \left(1 - \frac{1}{2^{2r+1}} + \frac{1}{3^{2r+1}} \cdots \right).$$

So that $\bar{E}(x)$ becomes

$$\bar{E}(x) = -\sum_{r:1}^{\infty}\left(\frac{e^4\beta}{2}\right)^{r-1}\frac{1}{\Gamma(r)}\left(1-\frac{1}{2^{2r-1}}+\frac{1}{3^{2r-1}}\cdots\right),$$

which is equal to $-E$; so for W' the following result appears:

$$W' = \frac{1}{\pi}\int_0^\infty \exp\left(-\frac{e^4\beta k}{2}\right)\frac{dk}{\sqrt{k}}\int_0^\infty \frac{e^t dt}{(e^t+1)^2} \times \cos\frac{t}{\sqrt{k}}\,2e^4 + 2\sum_{r:1}^{\infty}\left(\frac{e^4\beta}{2}\right)^{r-1}\frac{1}{\Gamma(r)} \times\left(1-\frac{1}{2^{2r-1}}+\frac{1}{3^{2r-1}}+\cdots\right)e^4 2.$$

After integration with respect to β one still has to determine the constant C appearing in

$$W-1 = -\frac{4}{\pi}\int_0^\infty \exp\left(-\frac{e^4\beta k}{2}\right)\frac{dk}{k^{3/2}}\int_0^\infty \frac{e^t dt}{(e^t+1)^2} \times\cos\frac{t}{k^{1/2}} + 4\cdot 2\sum_{r:1}^{\infty}\left(\frac{e^4\beta}{2}\right)^{r}\frac{1}{\Gamma(r+1)} \times\left(1-\frac{1}{2^{2r-1}}+\frac{1}{3^{2r-1}}+\cdots\right)+C.$$

To determine C the limit $\beta\to 0$ is taken:

$$C = \lim_{\beta\to 0}\frac{4}{\pi}\int_0^\infty \exp\left(-\frac{e^4\beta k}{2}\right)\frac{dk}{k^{3/2}}\int_0^\infty \frac{e^t dt}{(e^t+1)^2} \times\cos\frac{t}{\sqrt{k}}.$$

After interchanging the order of integration and introducing some elementary substitutions, one is faced with

$$C = \lim_{\beta\to 0}\int_{-\infty}^{+\infty}\frac{e^t dt}{(e^t+1)^2}\,\frac{2}{\pi}\int_{-\infty}^{+\infty}\exp\left(-\frac{e^4\beta}{2\tau^2}\right)e^{i\tau t}d\tau.$$

One finds for the value of C

$$C = 4\int_{-\infty}^{+\infty}\frac{e^t}{(e^t+1)^2}\,dt\,\delta(t) = 1,$$

so W becomes

$$W = -\frac{4}{\pi}\int_0^\infty \exp\left(-\frac{e^4\beta k}{2}\right)\frac{dk}{k^{3/2}}\int_0^\infty \frac{e^t dt}{(e^t+1)^2}\cos\frac{t}{\sqrt{k}} + 2\left[1+4\sum_{r:1}^{\infty}\left(\frac{e^4\beta}{2}\right)^{r}\frac{1}{\Gamma(r+1)} \times\left(1-\frac{1}{2^{2r-1}}\,\frac{1}{3^{2r-1}}\cdots\right)\right]. \tag{3.36}$$

The series development

$$d = \sum_{r:1}^{\infty}\left(\frac{e^4\beta}{2}\right)^{r}\frac{1}{\Gamma(r+1)}\sum_{s:1}^{\infty}(-1)^{s+1}\frac{1}{s^{2r-1}}$$

can be transformed to

$$d = \lim_{x\to 1}\sum_{s:1}^{\infty}(-1)^{s+1}sx^{s-1}\left[\exp\left(\frac{e^4\beta}{2s^2}\right)-1\right]$$

or

$$d = \lim_{x\to 1}\sum_{s:1}^{\infty}(-1)^{s+1}sx^{s-1}\exp\left(\frac{e^4\beta}{2s^2}\right) - \lim_{x\to 1}\sum_{s:1}^{\infty}(-1)^{s+1}sx^{s-1}$$

so that

$$d = \sum_{s:1}^{\infty}(-1)^{s+1}s\exp\left(\frac{e^4\beta}{2s^2}\right)-\frac{1}{4}$$

(where use has been made of Euler's summation proces for divergent series).

Using this result and (3.36), we have that the final expression for W is

$$W = \int d^3\mathbf{r}_\beta\int_{0/0}^{\mathbf{r}_\beta/\beta}\exp\left(-\tfrac{1}{2}\int_0^\beta\dot{\mathbf{r}}^2dt + e^2\int_0^\beta\frac{dt}{|r(t)|}\right)\mathfrak{D}\mathbf{r}(t) = -\frac{8}{\pi}\int_0^\infty\exp\left(-\frac{e^4\beta}{2}k^2\right)\frac{dk}{k^2}\int_0^\infty\frac{e^t dt}{(e^t+1)^2}\cos\frac{t}{k} + 8\sum_{n:1}^{\infty}(-1)^{n+1}n\exp\left(\frac{e^4\beta}{2n^2}\right). \tag{3.37}$$

In these equalities the first term of the last member contains only information about the continuous spectrum, while the series development is only related to the discrete spectrum.

The coefficient of β in the exponent of the second series, describing the discrete spectrum, provides the Bohr formula $E_n = -e^4/2n^2$ calculated by a straightforward path integral approach. In Appendices B and C it is shown how these path integral results can be transformed to reveal the s-type wavefunctions for both the bound states and the scattering states and even the corresponding density matrix. Our result for W cannot be compared directly with the Green's function in the momentum representation as obtained in Refs. 14, 16, and 17. Starting from the coordinate-space Green's function it is, in principle, possible to get (3.37).

4. DISCUSSION

Our final expression contains all information on the s-like energy levels and wavefunctions of the hydrogen atom. This holds true as well for the bound states as for the scattering states. Although after the integration over $\mathbf{r}_\beta$ several integrands for W might *a priori* lead to the same expression for W, the expression W (3.37) naturally transforms to the expressions at the end of Appendices B and C, whose integrands contain the hydrogen wavefunctions.

Of course the calculations in Appendices B and C merely serve for checking the expression for W (3.37) which is exact *a priori*. Strictly speaking, however, our expression W does not provide us with wavefunctions. In Ref. 11 we show how a slight modification of the method allows us to obtain the wavefunctions.

In the expression for W the contributions from the bound states and the scattering states are exactly separated. This separation has been facilitated by the application of Grosjean's theorem concerning the development of Laplace transforms in the neighborhood of the origin. The calculation of W presented above is a direct analytic calculation which is entirely self-consistent within the path integral formalism. It is, given the defining expression for the path integral and how it contains the energy and wavefunctions of our system, one arrives at the final expression for W and the energy levels without anything else than calculus.

Of course, the path integral has been transformed into an expansion of Riemann integrals containing only Gaussian path integrals, so that we did not calculate the hydrogen atom path integral by *actually* performing the integrations over paths and taking the appropriate limit as can be done for the harmonic oscillator.

The modified perturbation expansion converges more strongly than the development for the exponential function $\exp[|V(0)|]$ if $V(0)\neq 0$. The strong convergence of this expression becomes even more apparent since it also converges for the hydrogen atom, for which $V(0)=\infty$. A key feature of the modified expansion is that $\beta\neq\infty$ in our formalism, and there lies the difference with standard perturbation techniques of quantum theory.

Several of Feynman's ideas [like the introduction of the Fourier transform of the potential and defining $f(t)$] from his work on polarons[12] have been fruitful for the present work. Our own work on polarons[10] was also helpful because it faced us with a potential more difficult than the Coulomb potential and of which the Coulomb potential is a special case. The methods developed in Ref. 10 although perhaps not powerful enough to treat the polaron problem analytically at all α, turn out to be general enough to give the exact solution for the s states in the Coulomb potential case.

The introduction of the function W instead of $\rho(\mathbf{r}_0, \mathbf{r}_\beta)$ has provided important mathematical simplifications. If one treats $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$, or the trace, one has in addition the term $-t_it_j/\beta$ in the elements of $\det[B_n]$, Eq. (3.10). We do not see how in that case the full analytic treatment could be carried through. Integration over $\mathbf{r}_\beta$ to obtain W eliminates the time mixing in the determinant. This is the key feature which makes it possible to perform the integrations over the time variables. Another advantage of considering W instead of $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$ is that W can be approximated via a Laplace transform by a continued fraction method in some cases because the appropriate moments converge.

These moments diverge, in general, if $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$ is approximated. This provides us with an approximation method in the case where the summation cannot be done exactly. A disadvantage of considering W instead of $\rho(\mathbf{r}_0, \mathbf{r}_\beta)$ is the loss of information on a number of eigenstates (only s-like states are conserved for the hydrogen atom). This difficulty can be removed, however. Let us illustrate this point for the $2p$ states. Instead of calculating W, we calculate

$$W' = \int \mathbf{r}_\beta \cdot [\nabla_{\mathbf{r}_0} \rho(\mathbf{r}_\beta, \mathbf{r}_0)]_{\mathbf{r}_0=0}\, d^3\mathbf{r}_\beta .$$

This expression now has nonvanishing contributions from $2p$ states; the mathematics is only slightly more involved. A more important disadvantage of the method is that W does not give us, without further study, all the results of quantum statistical physics which derive from $\rho(\mathbf{r}_\beta, \mathbf{r}_0)$ in a standard way. For that purpose further study is necessary.

It may still be noted that the integral (3.18) is a generalization of a type of integrals treated for the first time by Grosjean.[15] This type of integral seems to be typical for Coulomb-type potentials and arises in an analogous form in polaron theory.[10,12]

5. CONCLUSION

We have calculated $W=\int d^3\mathbf{r}_\beta K(\mathbf{r}_\beta, 0)$, where $K(\mathbf{r}_\beta, 0)$ is the density matrix of the hydrogen atom. The calculation is direct, analytical, and entirely self-consistent within the path integral formalism of Feynman. To perform the calculations we have introduced a "modified perturbation method," which essentially consists in developing W in powers of $\sqrt{\beta}$. The energy spectrum for the s states of the hydrogen atom energies follows from this calculation in a straightforward way. It is shown how the other states can be included. It is also shown how W can be manipulated to obtain the wavefunctions, but this is by an indirect procedure. Summation of W provides us with an example where a problem of moments with divergent moments has been solved.

ACKNOWLEDGMENT

We are indebted to Professor C. C. Grosjean for the careful reading of the manuscript and interesting remarks.

Note: C. C. Grosjean ["An Apparent Inconsistency in Feynman's Path Integral Formalism of Nonrelativistic Quantum Mechanics," Proceedings of the Royal Flemish Academy of Sciences, Literature and Fine Arts of Belgium, Brussels, 1970. (to be published)] calculated the path integral for the case of a potential energy $V(r)$ which is spherically symmetric. He performed the $2(n-1)$ integrations over the spherical angular coordinates following Feynman's prescription. Grosjean did not consider any explicit form for $V(r)$. He also obtained the same propagator on the basis of another method which is based partly on Schrödinger's formalism and partly on Feynman's formalism. The two expressions he obtained this way contain remarkable differences before the limit of infinite time slicing is taken. Grosjean showed that both expressions, as they should, lead to the same limit.

APPENDIX A

The evaluation of the Gaussian integral

$$P_n = \int \exp\left[-\left(\tfrac{1}{2}\int_0^\beta \dot{\mathbf{r}}^2 dt + \int_0^\beta \mathbf{f}_n(t)\mathbf{r}(t)dt\right)\right]\mathfrak{D}\mathbf{r}(t), \tag{A1}$$

with initial conditions

$$t = 0, \quad \mathbf{r}(0) = \mathbf{r}_0, \quad t = \beta, \quad \mathbf{r}(\beta) = \mathbf{r}_\beta \tag{A2}$$

is explained.

The negative of the argument of the exponential function can be considered as the action of a particle in a potential

$$V(\mathbf{r}) = \mathbf{f}_n(t)\mathbf{r}(t). \tag{A3}$$

The classical path defined by this action, with initial conditions (A2) is easily seen to be determined by

$$\mathbf{r}_{cl,n} = -i\sum_{j:1}^{n}\mathbf{k}_j(t-t_j) + \dot{\mathbf{r}}(0)t + \mathbf{r}(0) \tag{A4}$$

or

$$\mathbf{r}_{cl,n} = i\sum_{j:1}^{n}\mathbf{k}_j[-(t-t_j)H(t-t_j) + (\beta - t_j)]\frac{t}{\beta} + \frac{t}{\beta}(\mathbf{r}_\beta - \mathbf{r}_0) + \mathbf{r}_0. \tag{A5}$$

After partial integration, the classical action can be written as

$$S_{cl,n} = -\tfrac{1}{2}\dot{\mathbf{r}}_{cl,n}\mathbf{r}_{cl,n}\Big|_0^\beta + \tfrac{1}{2}\int \mathbf{f}_n(t)\mathbf{r}_{cl,n}(t)dt. \tag{A6}$$

With the expression (A4) for $\mathbf{r}_{cl,n}$, this results in

$$S_{cl,n} = -\frac{1}{2\beta}(\mathbf{r}_\beta - \mathbf{r}_0)^2 + i\sum_{j:1}^{n}\mathbf{k}_j\frac{t_j}{\beta}(\mathbf{r}_\beta - \mathbf{r}_0) + i\mathbf{r}_0\sum_{j:1}^{n}\mathbf{k}_j - \tfrac{1}{2}\sum_{i:1}^{n}\sum_{j:1}^{n}\mathbf{k}_i\mathbf{k}_j \times\left[-(t_i - t_j)H(t_i - t_j)\;\left(1 - \frac{t_j}{\beta}\right)t_i\right]. \tag{A7}$$

Using the theorem of Feynman concerning Gaussian path integrals one gets as an expression for P_n

$$P_n = e^{S_{cl,n}}\int \exp\left(-\tfrac{1}{2}\int_0^\beta \dot{\mathbf{y}}^2 dt\right)\mathfrak{D}\mathbf{y}(t).$$

The remaining path integral corresponds to a free particle with initial conditions

$$t = 0, \quad \mathbf{y}(0) = 0, \quad t = \beta, \quad \mathbf{y}(\beta) = 0.$$

The final result for P_n is

$$P_n = e^{S_{cl,n}}\, 1/\sqrt{2\pi\beta}^3,$$

with

$$S_{cl,n} = -\frac{1}{2\beta}(\mathbf{r}_\beta - \mathbf{r}_0)^2 + i\sum_{j:1}^{n}\mathbf{k}_j\frac{t_j}{\beta}(\mathbf{r}_\beta - \mathbf{r}_0) + i\mathbf{r}_0\sum_{j:1}^{n}\mathbf{k}_j - \tfrac{1}{2}[k_1^2T'_{1,1} + \cdots + \mathbf{k}_n^2T'_{n,n} + 2\mathbf{k}_1\mathbf{k}_2T'_{1,2} + \cdots + 2\mathbf{k}_{n-1}\mathbf{k}_nT'_{n-1,n}],$$

where

$$T'_{i,j} = -t_it_j/\beta + \tfrac{1}{2}(t_i + t_j) - \tfrac{1}{2}|t_i - t_j|.$$

APPENDIX B

The transformation of the expression

$$B = 8\sum_{n:1}^{\infty}(-1)^{n+1}\,n\exp\left(\frac{e^4\beta}{2n^2}\right)$$

to Schrödinger notation is given.

First the expression

$$\frac{\partial B}{\partial\beta} = 8\sum_{n:1}^{\infty}(-1)^{n+1}\frac{e^4}{2n^2}\,n\exp\left(\frac{e^4\beta}{2n^2}\right)$$

is transformed.

One has

$$b = (-1)^{n+1}n2^4 = (-1)^{n+1}[(n-1)2^4 + 2^4].$$

b can also be written as

$$b = \frac{d^2}{ds^2}\frac{1}{s}\frac{s^n - (s-1)^n}{ns^{n-1}}\bigg|_{s\to 1/2}.$$

Introducing the Gaussian hypergeometric function, one has

$$b = \frac{d^2}{ds^2}\frac{1}{s}\,{}_2F_1\left(1, -n+1|2|\frac{1}{s}\right)\bigg|_{s=1/2}.$$

Use of the well-known Laplace transform

$$\frac{1}{s}\,{}_2F_1\left(1, -n+1|2|\frac{1}{s}\right) = \int_0^\infty e^{-\tau s}\,{}_1F_1(-n+1|2|\tau)d\tau$$

results in

$$b = \int_0^\infty \tau^2 e^{-\tau/2}\,{}_1F_1(-n+1|2|\tau)d\tau,$$

So $\partial B/\partial b$ can be written as

$$\frac{\partial B}{\partial b} = 2^2\sum_{n:1}^{\infty}\frac{e^4}{2n^2}\exp\left(\frac{e^4\beta}{2n^2}\right)\int\tau^2 e^{-\tau}\,{}_1F_1(-n+1|2|2\tau)d\tau$$

or

$$\frac{\partial B}{\partial b} = \int_0^{2\pi}d\varphi\int_0^\pi \sin\theta d\theta\int_0^\infty r^2 dr\sum_{n:1}^{\infty}e^{-\kappa r}(2\kappa)^3 \times\tfrac{1}{2}\,{}_1F_1(-n+1\;|2|2r)\;{}_1F_1(-n+1|2|0) \times Y_0^0(\theta,\varphi)\;Y_0^0(0,0)\exp\left(\frac{\beta e^4}{2n^2}\right)\frac{e^4}{2n^2},$$

with

$$\kappa = e^2/n,$$

After carrying out the integration with respect to β, one gets

$$8\sum_{n:1}^{\infty}(-1)^{n+1}\,n\exp\left(\frac{e^4\beta}{2n^2}\right) = C + \int d^3\mathbf{r}_\beta\sum_{n:1}^{\infty}e^{-\kappa|\mathbf{r}_\beta|}(2\kappa)^3\,\tfrac{1}{2} \times{}_1F_1(-n+1|2|2\kappa r_\beta)\;{}_1F_1(-n+1|2|0) \times Y_0^0(\theta,\varphi)Y_0^0(0,0)\exp(\beta e^4/2n^2).$$

To calculate the value of C the limit for $\beta \to 0$ is taken in both members of the last equality.

Using Euler's prescription to sum divergent series, one finds

$$2 = C + 2.$$

This results in

$$8 \sum_{n:1}^{\infty} (-1)^{n+1} n \exp\left(\frac{e^4\beta}{2n^2}\right) = \int d^3\mathbf{r}_\beta \sum_{n:1}^{\infty} e^{-\kappa r_\beta} (2\kappa)^3$$
$$\times \tfrac{1}{2}\, {}_1F_1(-n+1)|2|2\kappa r_\beta)\, {}_1F_1(-n+1)|2|0)$$
$$\times Y_0^0(\theta,\varphi) Y_0^0(0,0) \exp\left(\frac{\beta e^4}{2n^2}\right),$$

which gives the energy levels and wavefunctions for the bound s-like hydrogen states.

APPENDIX C

The formal transformation of the expression

$$S = -\frac{8}{\pi} \int_0^\infty \exp\left(-\frac{e^4\beta}{2} k^2\right) \frac{dk}{k^2} \int_0^\infty \frac{e^t dt}{(e^t+1)^2} \cos\frac{t}{k}$$

to Schrödinger notation is developed.

Using symmetry arguments one has

$$S = -\frac{4}{\pi} \int_0^\infty \exp\left(-\frac{e^4\beta}{2} k^2\right) \frac{dk}{k^2} \times \int_{-\infty}^{+\infty} \frac{e^t dt}{(e^t+1)^2} \cos\frac{t}{k}.$$

After a partial integration with respect to the variable t, one finds

$$S = \frac{4}{\pi i} \int_0^\infty \exp\left(-\frac{\beta}{2} k^2\right) \frac{dk}{k} \int_{-\infty}^{+\infty} \frac{e^t dt(1-e^t)}{(e^t+1)^3} e^{ie^2 t/k}.$$

After carrying out the substitution $e^t = z/(1-z)$, there appears

$$S = -\frac{4}{\pi} \int_0^\infty \exp\left(-\frac{\beta}{2} k^2\right) \frac{dk}{k} \int_{-1}^{0} \frac{dZ}{(1+2Z)^3} \times \exp\left(\frac{ie^2}{k} \log\frac{Z}{1+Z}\right).$$

Putting $Z = -t$, one obtains

$$S = \frac{4}{\pi} \int_0^\infty \exp\left(-\frac{\beta}{2} k^2\right) k^2 dk \frac{1}{(2ik)^3} \frac{\exp(e^2/k\pi)}{\beta(1+ie^2/k, 1-ie^2/k)} \left|\Gamma\left(1+\frac{ie^2}{k}\right)\right|^2 \int_0^1 \frac{t^{ie^2/k}(1-t)^{-ie^2/k}}{(\frac{1}{2}-t)^3}$$

or

$$S = \frac{4}{\pi} \int_0^\infty \exp\left(-\frac{\beta}{2} k^2\right) k^2 dk \frac{1}{(2ik)^3} \frac{d^2}{ds^2} \frac{\exp[(e^2/k)\pi] |\Gamma(1+ie^2/k)|^2}{\beta(1+ie^2/k, 1-ie^2/k)\, s} \int_0^1 \frac{t^{ie^2/k}(1-t)^{-ie^2/k}}{(1-t)^3} dt$$

Introducing the appropriate hypergeometric function this expression can be written as

$$S = \frac{2}{\pi} \int_0^\infty \exp\left(-\frac{\beta}{2} k^2\right) k^2 dk \frac{\exp(e^2\pi/k)}{(2ik)} \left|\Gamma\left(1+\frac{ie^2}{k}\right)\right|^2 \frac{d^2}{ds^2} \int_0^\infty d\tau e^{-s\tau} F\left(1+\frac{ie^2}{k}\middle|2\middle|\tau\right)\Bigg|_{s=1/2}$$

and further transformed into

$$S = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \int d^3\mathbf{r} \exp\left(\frac{\pi e^2}{k}\right) \exp(-ikr) \left|\Gamma\left(1+\frac{ie^2}{k}\right)\right|^2 F\left(1+\frac{ie^2}{k}\middle|2\middle|2ikr\right) \exp\left(-\frac{\beta k^2}{2}\right).$$

Because of the equivalence

$$\sum_k \leftrightarrows d^3\mathbf{k}/(2\pi)^3, \qquad V = 1,$$

one finally has

$$S = \sum_k \int d^3\mathbf{r} \exp(\pi e^2/k) |\Gamma[1+(ie^2/k)]|^2 \exp(-ikr) e^{i0\varphi} P_0^0(1) P_0^0(\cos\theta) F[1+(ie^2/k)|2|2ikr] F[1+ie^2/k)|2|0)$$
$$\times \exp(-\tfrac{1}{2}\beta k^2).$$

In this expression the wavefunctions for the s-like scattering states appear.

---

* In partial fulfilment of the requirements for a Ph.D.

† Also at the Solid State Physics Dept., Nuclear Energy Center, Mol Belgium.

[1] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[2] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[3] R. G. Storer, J. Math. Phys. **9**, 964 (1968).

[4] D. Gelman and L. Spruch, J. Math. Phys. **10**, 2240 (1969).

[5] M. Gutzwiller, J. Math. Phys. **8**, 1979 (1967).

[6] L. Schulman, Phys. Rev. **176**, 1558 (1968).

[7] A. Messiah, *Mécanique Quantique* (Dunod, Paris, 1959), p. 202.

[8] C. Garrod, Rev. Mod. Phys. **38**, 483 (1966).

[9] C. C. Grosjean, Bull. Soc. Mat. Belg **18**, 251 (1965).

[10] M. Goovaerts, A. Babcenco, and J. Devreese, "A New Expansion Method in the Feynman Path Integral Formalism. Application to a One-Dimensional δ-Function Potential," J. Math. Phys. (to be published).

[11] J. Devreese and M. Goovaerts, "A Modified Perturbation

Development for Energy States. Application to the Calculation of Self-Energy, and Effective Mass of an Electron in a Polar Crystal," (unpublished).
[12] C. G. Kuper and G. D. Whitfield, *Polarons and Excitons* (Oliver and Boyd, Edinburgh, 1962).
[13] M. Goovaerts, thesis (Ghent, 1971)(unpublished).
[14] Hostler, J. Math. Phys. **5**, 1235 (1964).
[15] C. C. Grosjean, Simon Stevin, **41** (1967).
[16] S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960).
[17] J. Schwinger, J. Math. Phys. **5**, 1606 (1964).