On the coupling of self-conjugate systems with $\overline{SL(3,R)}$ symmetry

J.A. Castilho Alcarás [a)]

Instituto de Fisica Teórica, São Paulo, Brazil

L. C. Biedenharn[b)]

Institut für Theoretische Physik der Johann Wolfgang Goethe Universität, Frankfurt/Main, Federal Republic of Germany

(Received 28 September 1978)

The coupling of unitary self-conjugate $\overline{SL(3,R)}$ multiplicity-free irreps is explicitly calculated. It is proven that pair-wise coupling of a finite number of these self-conjugate irreps never contains a continuum $\overline{SL(3,R)}$ irrep which is multiplicity-free. The significance of $\overline{SL(3,R)}$ symmetry in nuclear and hadronic systems is discussed, including implications for the coupling of quarkels.

1. ORIGIN AND SIGNIFICANCE OF THE PROBLEM

Systems having $\overline{SL(3,R)}$ symmetry are of fundamental importance in both nuclear structure physics[1-5] and hadronic structure ("quarks") physics.[1,6-8] The elementary structures (unitary irreps) of $\overline{SL(3,R)}$ are all known,[9] and among these there are precisely three self-conjugate multiplicity-free irreps (not counting the identity) which are of particular interest since they are *discrete*, and might possibly serve as models for fundamental "particles." Of these three self-conjugate multiplicity-free irreps, there is a unique spinorial irrep which is of special interest since it can be considered as a model for a "deformable quark."[10]

The present paper has as its purpose the investigation of all possible couplings of these self-conjugate multiplicity-free irreps among themselves, preserving $\overline{SL(3,R)}$ symmetry and ascribing to these irreps both bosonic and fermionic symmetry character. These results—which are discussed in detail in Sec. 2—have one surprising aspect: *it is not possible to couple, pairwise, any finite number of self-conjugate multiplicity-free irreps to produce a multiplicity-free irrep lying in the continuum.*[11] We prove this by direct construction in Sec. 2 and by a more elegant argument in Sec. 3.

Let us indicate briefly the importance of $\overline{SL(3,R)}$ in nuclear structure physics. The group $SL(3,R)$, is, by definition, the group of rotations and volume-preserving deformations in three-dimensional space. Since it is known[12] that nuclear forces tend strongly to conserve nuclear volume, but no nuclear shape, the group $SL(3,R)$ should accordingly be a useful approximate description of rotations and vibrations (below the onset of breathing modes, say). (This view was originally suggested in Ref. 1 and applied in Ref. 2.) Since the group $SL(3,R)$ is noncompact, this approximate description of nuclear structure leads to rotational bands that are *non*terminating, in sharp contrast to the terminating band structure of the Elliot model [SU(3)]. In a sense both $SL(3,R)$ and SU(3) deal with the same (shell model) harmonic oscillator structure, but organize the levels differently. In $SL(3,R)$ *all* levels (in the three-dimensional harmonic oscillator) having a given angular momentum are organized into coherent (continuum) states. Unlike Elliot SU(3), which is its own covering group, the covering $\overline{SL(3,R)}$ (double covering) can, however, possess spinors. This makes possible models in which spin can be viewed as *intrinsic*.

Extensions of the $\overline{SL(3,R)}$ symmetry to include volume changes have been discussed in Refs. 3 and 4 (for nuclear physics).

It would appear from this discussion that $\overline{SL(3,R)}$ symmetry is heuristic, and not fundamental to nuclear structure. It has been shown recently,[5] however, that the collective motion of any (nonrelativistic) many-particle system necessarily includes, besides the angular momentum, a vortex angular momentum which together with the three intrinsic collective momenta generate $GL(3,R) \supset SL(3,R)$. It follows that these symmetries are indeed more fundamental than hitherto expected; whether or not the nuclear spectra actually display this symmetry becomes a particularity of specific nuclear forces.

The introduction of $SL(3,R)$ as a fundamental symmetry of *hadronic matter* was first proposed in Ref. 1. The motivation for this proposal lies in the Regge band structure where the hadrons (having specified parity and flavor quantum numbers) are found to lie on approximately linear trajectories ($M^2 \propto$ spin) having spin $J, J+2, J+4, \cdots$. Irreps of $SL(3,R)$ with such band structure were given in Ref. 1; the generalization to $\overline{SL(3,R)}$ was given in Ref. 13. The unique self-conjugate multiplicity-free spinor irrep (the *quarkel*) was first found by Joseph,[14] and independently by Weaver.[15,16] The proper embedding of the quarkel as a quantal Regge band in a fully explicit relativistic Poincaré-covariant structure was demonstrated in Ref. 17.

[a)]Supported in part by the National Science Foundation (U.S.A.) for a semester at Duke University.

[b)]Alexander von Humboldt Foundation, Senior U.S. Scientist Award, on leave from Duke University, Durham, North Carolina, 27706.

 0022-2488/79/081615-07$01.00

The importance of the unique spinorial irrep for mathematical physics was only recently pointed out in a paper by Ne'eman[7] (correcting, incidentally, a misinterpretation of a result of Cartan). For the generalization of the Poincaré relativistic symmetry to affine relativity[8] the quarkel irrep serves as a *global* spinor.

Despite the brevity of the above sketch, it should be clear that $\overline{\mathrm{SL}(3,R)}$, and the quarkel in particular, have a valid claim of being fundamental. If this is accepted—and the occurrence of $\overline{\mathrm{SL}(3,R)}$ symmetry in so many distinct branches of physics supports this view—then it is of importance to investigate in detail how to implement $\overline{\mathrm{SL}(3,R)}$ symmetry-preserving couplings. It is natural to consider, for example, models in which a collection of fermion quarkels interact to produce a sort of "fluid," very much like a shell structure in nuclear physics but in which (unlike, say, the $f_{7/2}$ shell) the shell structure can accommodate *unlimitedly* many fermions (see Sec. 3).

The first step in such a program is to determine the possible couplings which preserve the symmetry. This is accomplished in the next section.

2. EXPLICIT CALCULATIONS

A. Resumé of the Lie algebra of SL(3,R)

Both SL(3,R) and its covering group $\overline{\mathrm{SL}(3,R)}$ have eight infinitesimal generators, usually chosen in such a way that three of them are generators of SO(3), the maximal compact subgroup of SL(3,R), while the remaining five are the components of a quadrupole operator. Written in a normalized spherical basis, the commutation relations satisfied by these generators are:

$$
\begin{aligned}
&[J_i,J_{\lambda 1}] = \pm J_{\pm 1}, \quad [J_{+1},J_{-1}] = -J_0,\\
&[J_{\pm 1},T_m] = \mp[3-\tfrac{1}{2}m(m+1)]^{1/2}T_{m\pm 1},\\
&m=0,\pm 1,\pm 2, \quad [T_{+2},T_{-2}] = -4J_0.
\end{aligned}
\tag{2.1}
$$

These commutation relations given above constitute a minimal set of commutation relations; the remaining ones are obtained combining properly Eqs. (2.1). In doing so, one obtains the formulas

$$
\begin{aligned}
&[J_0,T_m] = mT_m, \quad m=0,\pm 1,\pm 2,\\
&[T_m,T_{m'}] = -\sqrt{10}\,\langle 2m2m'|1m+m'\rangle J_{m+m'}.
\end{aligned}
\tag{2.2}
$$

Defining

$$
\begin{aligned}
T^2 &= \sum_m (-)^m T_m T_{-m},\\
JJT &= -\sqrt{3/35}\sum_{m_1,m_2}(-)^{m_1+m_2}\\
&\quad\times\langle 1m_1 1m_2|2m_1+m_2\rangle J_{m_1}J_{m_2}T_{-m_1-m_2},\\
T^3 &= \frac{3}{\sqrt{5}}\sum_{m_1,m_2}(-)^{m_1+m_2}\\
&\quad\times\langle 2m_1 2m_2|2m_1+m_2\rangle T_{m_1}T_{m_2}T_{-m_1-m_2},
\end{aligned}
\tag{2.3}
$$

it can be verified that the operators

$$
\mathscr{I}_2 = \tfrac{1}{12}(J^2-\tfrac{1}{2}T^2), \quad \mathscr{I}_3 = \frac{i}{24}(JJT+\tfrac{1}{3}T^3),
\tag{2.4}
$$

are Casimir invariants of SL(3,R).

To find irreps of $\overline{\mathrm{SL}(3,R)}$ in the chain $\overline{\mathrm{SL}(3,R)} \supset \mathrm{SU}(2) \supset \mathrm{U}(1)$, one first writes the basis states as $|p,q;Jm\rangle$, where p,q labels the irrep of $\overline{\mathrm{SL}(3,R)}$ one is dealing with while J and M are internal quantum labels related to the angular momentum and its z projection. The matrix elements of the operators J_m have their usual expressions. For the matrix elements of the T_m operators one exploits their irreducible tensorial character relative to SU(2) to write

$$
\begin{aligned}
&\langle p,q;J'M|T_m|p,q;JM\rangle\\
&\quad = \frac{\langle p,q;J'\|T\|p,q;J\rangle}{\sqrt{2J'+1}}\,\langle JM\,2m|J'm'\rangle.
\end{aligned}
\tag{2.5}
$$

Now, using Eqs. (2.2), the recoupling relation[18]

$$
\begin{aligned}
&\langle a\alpha b\beta|e\alpha+\beta\rangle\langle e\alpha+\beta d\delta|c\gamma\rangle\\
&\quad = \sum_f \sqrt{(2e12)(2f+1)}\langle b\beta d\delta|f\beta+\delta\rangle\\
&\qquad\times\langle a\alpha f\beta+\delta|c\gamma\rangle W(abcd;ef),
\end{aligned}
\tag{2.6}
$$

and the unitarity condition

$$
\langle p,q;J'\|T\|p,q;J\rangle = (-)^{J'-J}\langle p,q;J\|T\|p,q;J'\rangle,
\tag{2.7}
$$

one finds a set of equations relating the reduced matrix elements $\langle p,q;J\|T\|p,q;J'\rangle$. These equations can easily be solved in the case of irreps such that the matrix elements with $|J'-J|=1$ vanish identically. For those irreps one obtains[2]:

$$
\langle p,q;J\|T\|p,q,J\rangle = \eta\left(\frac{(2J)(2J+1)(2J+2)}{6(2J-1)(2J+3)}\right)^{1/2},
\tag{2.8}
$$

$$
\begin{aligned}
&|\langle p,q;J\|T\|p,q;J+2\rangle|^2\\
&\quad = \tfrac{1}{4}(2J)(2J-1)(2J-2)\left(1+\frac{\eta^2}{(2J-1)^2}\right),
\end{aligned}
\tag{2.9}
$$

where η is any real number and J can assume the values contained in any one of the following three sets:

$$
\begin{aligned}
&(\alpha)\colon J = \tfrac{1}{2},\tfrac{5}{2},\tfrac{9}{2},\cdots \text{ (with } \eta=0 \text{ only)},\\
&(\beta)\colon J = 0,2,4,\cdots,\\
&(\gamma)\colon J = 1,3,5,\cdots.
\end{aligned}
\tag{2.10}
$$

Since we have chosen $\langle J\pm 1\|T\|J\rangle = 0$, Eqs. (2.10) tell us that one has found three irreps. From (2.8) and (2.9) it follows that for these irreps the Casimir invariants (2.4) assume the values

$$
\mathscr{I}_2 = -\frac{1}{4}-\frac{\eta^2}{36}, \quad \mathscr{I}_3 = \frac{\eta^3 i}{9\sqrt{105}}.
\tag{2.11}
$$

Since under conjugation η goes to $-\eta$, the *discrete* irreps (2.10) with $\eta=0$ are self-conjugate. The three self-conjugate irreps have the same values for the invariants ($\mathscr{I}_2=\frac{1}{4}, \mathscr{I}_3=0$) and are distinguished by the labels (α), (β), (γ) denoting the spin content.[19] For this reason the states of these irreps will be denoted in the following by $|\eta,\mu;JM\rangle$, where μ can be (α) with $\eta=0$, or (β), (γ), with $\eta=0, \neq 0$.

B. Realization of the $\overline{SL(3,R)}$ generators by the Jordan–Schwinger map

Let b_{JM} be the fermion (or boson) creation operator of the state $|0,\mu;JM\rangle$ of any of the irreps (2.10) with $\eta = 0$ and $\bar{b}_{JM}$ the corresponding fermion (or boson) annihilation operator. It is well known that the one-body operators

$$\mathscr{X}_i = \sum_{JMJ'M'} \langle 0,\mu;J'M'|X_i|0,\mu;JM\rangle b_{J'M'}\bar{b}_{JM} \tag{2.12}$$

satisfy the same commutation relations as the generators X_i themselves ("Jordan–Schwinger mapping"). Therefore, the one-body operators

$$\mathscr{J}_m = \sum_{JM} \langle 0,\mu;JM+m|J_m|0,\mu;JM\rangle b_{JM+m}\bar{b}_{JM}, \quad m = 0, \pm 1, \tag{2.13}$$

$$\mathscr{T}_m = \sum_{JMJ'} \langle 0,\mu;J'M+m|T_m|0,\mu;JM\rangle b_{J'M}+m\bar{b}_{JM}, \quad m = 0, \pm 1, \pm 2, \tag{2.14}$$

provide a realization[20] of the $\overline{SL(3,R)}$ generators.

We now try to realize a generic state $|\eta,\mu;JM\rangle$ of any irrep of type (2.10) as a two-body operator acting on the vaccum, i.e.,

$$|\eta,\mu;JM\rangle = \mathscr{O}^{(\eta,\mu)}_{JM}|0\rangle. \tag{2.15}$$

Taking $\mathscr{O}^{(\eta,\mu)}_{JM}$ to be

$$\mathscr{O}^{(\eta,\mu)}_{JM} = \sum_{J_1M_1J_2M_2} A^{(J,M)}_{J_1M_1J_2M_2} b_{J_1M_1}b_{J_2M_2}, \tag{2.16}$$

one sees that in order that the realizations X_i of the generators have in the basis (2.15) the same matrix elements as the X_i in an abstract basis $|\eta,\mu;JM\rangle$, it is necessary and sufficient that the following commutation relations hold

$$[\mathscr{X}_i,\mathscr{O}^{(\eta,\mu)}_{JM}] = \sum_{J'M'} \langle \eta,\mu;J'M'|X_i|\eta,\mu;JM\rangle \mathscr{O}^{(\eta,\mu)}_{J'M'}. \tag{2.17}$$

By angular momentum theory, it follows that the operators

$$\mathscr{O}^{(\eta,\mu)}_{JM} = \sum_{J_1J_2} f^J_{J_1J_2} \sum_{M_1M_2} \langle J_1M_1J_2M_2|JM\rangle b_{J_1M_1}b_{J_2M_2} \tag{2.18}$$

satisfy Eq. (2.17) for $\mathscr{X}_i = \mathscr{J}_{+1}, \mathscr{J}_0, \mathscr{J}_{-1}$, irrespective of the values of the constants f. For $\mathscr{X}_i = \mathscr{T}_m$ $(m = 0, \pm 1, \pm 2)$, Eq. (2.17) imposes some constraints on the constants f.

In order to find those constraints, one starts by first evaluating the commutator of $\mathscr{T}_m$ by $\mathscr{O}^{(\eta,\mu)}_{JM}$. Using Eqs. (2.14), (2.18), (2.6), and (2.7) one finds

$$\begin{aligned}
&[\mathscr{T}_m,\mathscr{O}^{(\eta,\mu)}_{JM}] \sum_{J_1M_1J_2M_2f} (-)^{J_1+J_2+J}\langle J_1M_1J_2M_2|fM+m\rangle\langle JM\,2m|\,fM+m\rangle\sqrt{(2f+)} \\
&\times\Bigg(f^J_{J_1-2J_2})[(2J_1-2)(2J_1-1)(2J_1)]^{1/2}\begin{Bmatrix} J_2 & J & J_1-2 \\ 2 & J_1 & f \end{Bmatrix} + f^J_{J_1+2J_2}[(2J_1+2) \\
&\times(2J_1+3)(2J_1+4)]^{1/2}\begin{Bmatrix} J_2 & J & J_1+2 \\ 2 & J_1 & f \end{Bmatrix}\Bigg) b_{J_1M_1}b_{J_2M_2}.
\end{aligned} \tag{2.19}$$

Equating this expression to the RHS of Eq. (2.17) and using the linear independence of the set $\{b_{J_1M_1}b_{J_2M_2}$ with $J_1 \neq J_2$; $b_{J_1M_1}b_{J_1M_2}$ with $M_1 \neq M_2\}$ one obtains

$$\begin{aligned}
&f^J_{J_1-2J_2}[(2J_1-2)(2J_1-1)(2J_1)]^{1/2}\begin{Bmatrix} J_2 & J & J_1-2 \\ 2 & J_1 & J+\alpha \end{Bmatrix} f^J_{J_1+2J_2}[(2J_1+2)(2J_1+3)(2J_1+4)]^{1/2} \\
&\quad\times\begin{Bmatrix} J_2 & J & J_1+2 \\ 2 & J_1 & J+\alpha \end{Bmatrix} + (-)^{\alpha}+f^J_{J_1J_2-2}[(2J_2-2)(2J_2-1)(2J_2)]^{1/2}\begin{Bmatrix} J_1 & J & J_2-2 \\ 2 & J_2 & J+\alpha \end{Bmatrix} \\
&\quad+(-)^{\alpha}f^J_{J_1J_2+2}[(2J_2+2)(2J_2+3)(2J_2+4)]^{1/2}\begin{Bmatrix} J_1 & J & J_2+2 \\ 2 & J_2 & J+\alpha \end{Bmatrix} \\
&= (-)^{J_1+J_2+J}A(J_1,J_2,J,\alpha), \quad \text{for } \alpha = 0, \pm 1, \pm 2,
\end{aligned} \tag{2.20}$$

where

$$A(J_1,J_2,J,2) = \left(\frac{(2J+2)(2J+3)(2J+4)}{(2J+1)(2J+5)}\right)^{1/2}\left(1+\frac{\eta^2}{(2J+3)^2}\right)^{1/2}\Delta(J_1J_2J+2)f^{J+2}_{J_1J_2},$$

$$A(J_1,J_2,J,-2) = \left(\frac{(2J-2)(2J-1)(2J)}{(2J+3)(2J-3)}\right)^{1/2}\left(1+\frac{\eta^2}{(2J-1)^2}\right)^{1/2}\Delta(J_1J_2J-2)f^{J-2}_{J_1J_2},$$

$$A(J_1,J_2,J,0)=\left(\frac{(2J)(2J+2)(2J)}{3(2J-1)(2J+1)(2J+2)}\right)^{1/2}\eta\Delta(J_1J_2J)f^J_{J_1J_2},$$
$$A(J_1,J_2,J,\pm 1)=0, \tag{2.21}$$

and $\Delta(J_1J_2J_3)$ is equal to 1 when the three angular momenta J_1, J_2, and J_3 satisfy the triangularity relations, being zero otherwise.

The next step is to substitute in Eqs. (2.20) the explicit expressions of the $6-j$ symbols and try to solve the resulting set of equations in the unknown f's. Before doing that let us briefly examine which irreps of type (2.10) are expected to be realized in this way.

Since J_1 and J_2 are both integers or half-integers, J can assume only integer values; this excludes the irreps (2.10) of type (α).

When the b's are fermion operators, Eq. (2.18) implies that $f^J_{J_1J_2}=-(-)^{J_1+J_2-J}f^J_{J_2J_1}$ and, consequently $f^1_{1/2\,1/2}=0$ which, in turn, implies [by Eqs. (2.20)] that $\mathcal{O}^{(\eta,\gamma)}_{JM}$ with odd values of J also vanish identically, thereby excluding irreps of type (γ). Concerning irreps of type (β), only those with $\eta=0$ are realizable as can be seen by the arguments presented in the next subsection.

When the b's are boson operators, one has $f^J_{J_1J_2}=(-)^{J_1+J_2-J}f^J_{J_2J_1}$ and consequently $f^1_{J_1J_1}=0$ which implies, as seen above, that the $\mathcal{O}^{(\eta,\gamma)}_{JM}$ of the irreps of type (γ) vanish identically. For the irreps of type (β) one has the same situation as that when the b's were fermion operators.

Since not all irreps of type (2.10) can be realized by (2.18), it is instructive to examine which irreps can be obtained by coupling two kinematically independent self-conjugate (i.e., $\eta=0$) irreps (2.10).

C. The reduction of the direct product of two self-conjugate multiplicity-free irreps of $\overline{SL(3,R)}$

Let us consider the linear combinations of basis states of two self-conjugate irreps of type (2.10):

$$|\eta,\mu_3;JM\rangle=\sum_{J_1M_1J_2M_2}f^J_{J_1J_2}\langle J_1M_1J_2M_2|JM\rangle|0,\mu;J_1M_1\rangle|0,\mu_2;J_2M_2\rangle, \tag{2.22}$$

such that under the action of the generators of $\overline{SL(3,R)}$ they transform like the basis states of an irrep $[\eta,\mu_3]$ of type (2.10). That is, they must satisfy:

$$(J_m)_{\text{total}}|\eta,\mu_3;JM\rangle=\langle\eta,\psi_3;JM+m|J_m|\eta,\mu_3;JM\rangle|\eta,\mu_3;JM+m\rangle, \tag{2.23}$$

$$(T_m)_{\text{total}}|\eta,\mu_3;JM\rangle=\sum_{=0,\pm 2}\langle\eta,\mu_3;J+\alpha\ M+m|T_m|\eta,\mu_3;JM\rangle|\eta,\mu_3;JM\rangle|\eta,\mu_3;J+\alpha\ M+m\rangle, \tag{2.24}$$

where $(X)_{\text{total}}=X_{(1)}\otimes\mathbb{1}_{(2)}+\mathbb{1}_{(1)}\otimes X_{(2)}$, as usual. Since the Clebsch–Gordan coefficient in (2.22) vanishes if the J's do not satisfy the triangularity conditions, one may assume, without losing generality, that $f^J_{J_1J_2}=0$ when $\Delta(J_1J_2J_3)=0$.

Equation (2.23) is automatically satisfied due to the $(J_1)\otimes(J_2)\to(J)$ coupling existing in the definition (2.22). On the other hand, Eq. (2.24), by the use of Eqs. (2.5) and (2.6), implies that the f's must satisfy Eqs. (2.20) but now with no symmetry condition since $|0,\mu_1;JM_1M_1\rangle$ and $|0,\mu_2;J_2M_2\rangle$ are kinematically independent.

The structure of Eqs. (2.20) allows one to prove that only the irreps $[0,\mu_3]$ can be obtained by Eq. (2.22); i.e., only the irreps with $\eta=0$ are present in the multiplicity-free part of the reduction of $[0,\mu_1]\otimes[0,\mu_2]$. In order to prove this claim one has to examine the nine possibilities (see Table I) that one would expect for μ_3 in Eq. (2.22) by considering only the angular momentum content of $[0,\mu_1]$ and $[0,\mu_2]$.

Consider $\mu_3=(\beta)$. To obtain $J=0$ one must have $J_1=J_2$ which implies $\mu_1=\mu_2$. Taking $J=0$ and $\alpha=2$ in Eqs. (2.20), the LHS vanishes by the triangularity conditions while the RHS gives $f^2_{J_1J_1}$ multiplied by a nonzero number. Therefore, $f^2_{J_1J_1}$ must be zero. Now take $J=2$, $J_2=J_1\pm 2$, and $\alpha=0$ in Eq. (2.20). Two of the f's in the LHS vanish by triangularity while the other two vanish because they are of the form $f^2_{J_1J_1}$. On the other hand, for $\eta\neq 0$, the RHS of this equation gives $f^2_{J_1J_1\pm 2}$ multiplied by nonzero number, which implies $f^2_{J_1J_1\pm 2}=0$. Since $\mu_1=\mu_2$, the only J_2's that can be coupled to J_1 to give $J=2$ are $J_2=J_1,J_1\pm 2$. One concludes then that $|\eta\neq 0,(\beta);00\rangle$ given by Eq. (2.22) vanishes identically and, by Eqs. (2.20), the same occurs for all $|\eta\neq 0,(\beta);JM\rangle$. This proves that our claim is true for the possibilities 1, 5, and 7 of Table I. Similar arguments can be used to show that the claim is true for the remaining possibilities.

Now that we have proven that the irreps $[\eta\neq 0,\mu]$ are not present in the reduction of $[0,\mu_1]\otimes[0,\mu_2]$, let us find the coefficients f in Eq. (2.22) that make the reduction $[0,\mu_1]\otimes[0,\mu_2]\to[0,\mu_3]$. Substituting in Eqs. (2.20) the explicit expressions of the $6-j$ symbols[21] and putting $\eta=0$, one obtains the following set of equations:

$$f^J_{J_1\ \ 2\ J}\left(\frac{s(s+1)(s+2)(s+3)(u-1)u(u+1)(u+2)}{(2J_1-3)(2J_1+1)}\right)^{1/2}+f^J_{J_1+2\,J_2}\left(\frac{(v-1)v(v+1)(v+2)(t-1)t(t+1)(t+2)}{(2J_1+1(2J_1+5)}\right)^{1/2}$$

$$+f^{J}_{J_1J_2-2}\left(\frac{s(s+1)(s+2)(s+3)(t-1)t(t+1)(t+2)}{(2J_2-3)(2J_2+1)}\right)^{1/2}+f^{J}_{J_1J_2+2}\left(\frac{(v-1)v(v+1)(v+2)(u-1)u(u+1)(u+2)}{(2J_2+1)(2J_2+5)}\right)^{1/2}$$

$$=(2J+2)(2J+3)(2J+4)f^{J+2}_{J_1J_2} \text{ only for } \Delta(J_1J_2J+2)=1; \tag{2.25}$$

$$-f^{J}_{J_1-2\,J_2}\left(\frac{s(s+1)(s+2)(t+2)(u-1)u(u+1)(v-1)}{(2J_1-3)(2J_1+1)}\right)^{1/2}+f^{J}_{J_1+2\,J_2}\left(\frac{(s+3)(t-1)t(t+1)(u+2)v(v+1)(v+2)}{(2J_1+1)(2J_1+5)}\right)^{1/2}$$

$$+f^{J}_{J_1J_2-2}\left(\frac{s(s+1)(s+2)(t-1)t(t+1)(u+2)(v-1)}{(2J_2-3)(2J_2+1)}\right)^{1/2}-f^{J}_{J_1J_2+2}\left(\frac{(s+3)(t+2)(u-1)u(u+1)v(v+1)(v+2)}{(2J_2+1)(2J_2+5)}\right)^{1/2}$$

$$=0 \quad \text{only for } \Delta(J_1J_2J+1)=1; \tag{2.26}$$

$$f^{J}_{J_1-2\,J_2}\left(\frac{s(s+1)(t+1)(t+2)(u-1)u(v-1)v}{(2J_1-3)(2J_1+1)}\right)^{1/2}+f^{J}_{J_1+2\,J_2}\left(\frac{(s+2)(s+3)(t-1)t(u+1)(u+2)(v+1)(v+2)}{(2J_1+5)(2J_1+1)}\right)^{1/2}$$

$$+f^{J}_{J_1J_2-2}\left(\frac{s(s+1)(t-1)t(u+1)(u+2)(v-1)v}{(2J_2-3)(2J_2+1)}\right)^{1/2}+f^{J}_{J_1J_2+2}\left(\frac{(s+2)(s+3)(t+1)(t+2)(u-1)u(v+1)(v+2)}{(2J_2+5)(2J_2+1)}\right)^{1/2}$$

$$=0 \text{ only for } \Delta(J_1J_2J)=1; \tag{2.27}$$

$$-f^{J}_{J_1-2J_2}\left(\frac{st(t+1)(t+2)(u-1)(v-1)v(v+1)}{(2J_1-3)(2J_1+1)}\right)^{1/2}+f^{J}_{J_1+2J_2}\left(\frac{(s+1)(s+2)(s+3)(t-1)u(u+1)(u+2)(v+2)}{(2J_1+5)(2J_1+1)}\right)^{1/2}$$

$$+f^{J}_{J_1J_2-2}\left(\frac{s(t-1)u(u+1)(u+2)(v-1)v(v+1)}{(2J_2-3)(2J_2-1)}\right)^{1/2}$$

$$-f^{J}_{J_1J_2+2}\left(\frac{(s+1)(s+2)(s+3)t(t+1)(t+2)(u-1)(v+2)}{(2J_2+5)(2J_2+1)}\right)^{1/2}=0 \text{ only for } \Delta(J_1J_2J-1)=1; \tag{2.28}$$

$$f^{J}_{J_1-2\,J_2}\left(\frac{(t-1)t(t+1)(t+2)(v-1)v(v+1)(v+2)}{(2J_1-3)(2J_1+1)}\right)^{1/2}+f^{J}_{J_1+2\,J_2}\left(\frac{s(s+1)(s+2)(s+3)(u-1)u(u+1)(u+2)}{(2J_1+5)(2J_1+1)}\right)^{1/2}$$

$$+f^{J}_{J_1J_2-2}\left(\frac{(u-1)u(u+1)(u+2)(v-1)v(v+1)(v+2)}{(2J_2-3)(2J_2+1)}\right)^{1/2}+f^{J}_{J_1J_2+2}\left(\frac{s(s+1)(s+2)(s+3)(t-1)t(t+1)(t+2)}{(2J_2+5)(2J_2+1)}\right)^{1/2}$$

$$=(2J-2)(2J-1)(2J)f^{J-2}_{J_1J_2} \text{ only for } \Delta(J_1J_2J-2)=1; \tag{2.29}$$

where

$$s=J_1+J_2+J, \quad t=-J_1+J_2+J, \quad u=J_1-J_2+J, \quad v=J_1+J_2-J. \tag{2.30}$$

If one fixes μ_1, μ_2, and μ_3, these equations allow one to determine, up to a multiplicative constant, all the $f^{J}_{J_1J_2}$ for $J=J_0,J_0+2,J_0+4,\cdots$, where J_0 is the minimum value of J. For instance, when $\mu_1=\mu_2$ and $\mu_3=(\beta)$, one obtains

$$f^{0}_{JJ}=\lambda(-)^{(J-J_0)/2}\frac{(2J-2)!^4}{(2J)!^4}(2J+1)^{1/2}, \quad f^{2}_{JJ}=0,$$

$$f^{2}_{JJ+2}=f^{2}_{J+2\,J}=\lambda(-)^{(J-J_0)/2}\frac{(2J-2)!^4}{(2J)!^4}\left(\frac{(2J+2)(2J+3)}{6(2J+4)}\right)^{1/2},$$

$$f^{4}_{JJ}=\lambda(-)^{(J-J_0)/2}\frac{2J-2)!^4}{(2J)!^4}\left(\frac{(2J+3)(2J+2)(2J+1)(2J)(2J-1)}{(2J+5)(2J+4)(2J-3)(2J-2)}\right)^{1/2}, \quad f^{4}_{JJ+2}=f^{4}_{J+2\,J}=0,$$

$$f^{4}_{JJ+4}=f^{4}_{J+4\,J}=\lambda(-)^{(J-J_0)/2}\frac{2J-2)!^4}{(2J)!^4}\left(\frac{5(2J+7)(2J+6)(2J+3)(2J+2)}{56(2J+8)(2J+5)(2J+4)}\right)^{1/2}, \tag{2.31}$$

where[22]

$$n!^4=\begin{cases} n(n-4)(n-8),\ldots\begin{cases}4\\3\\2\\1\end{cases}, & \text{for } n=1,2,3,\cdots,\\ 1 & \text{for } n=0,-1,-2,\cdots\end{cases} \tag{2.32}$$

Obtaining all the $f^{J}_{J_1J_2}$ for the first values of J, for the eight possibilities of Table I, one observes that all of them fit into the general formula:

$$f^{J}_{J_1J_2}=\lambda(-)^{(v-2J_0)/4}\left(\frac{(s-1)!^4(s-2)!^4(t-3)!^4(t-2)!^4(u-3)!^4(u-2)!^4(v-2)!^4(v-3)!^4(2J_1+1)(2J_2+1)}{s!^4(s+1)!^4(t-1)!^4t!^4(u-1)!^4(v-1)!^4v!^4}\right)^{1/2} \tag{2.33}$$

for $(v - 2J_0)/4 =$ integer,

$f^J_{J_1J_2} = 0$, for $(v - 2J_0)/4 =$ not integer,

where J_0 is the minimum value of J in $[0,\mu_3]$.

The case $[0,(\beta)] \otimes [0,(\gamma)] \to [0,(\gamma)]$, although expected by angular momentum coupling, is *not* realized. When one solves Eqs. (2.25)–(2.29) for $J = 1$ and 3 one finds $f^1_{JJ+1} = f^1_{J+1J} = 0$. Since, in this case, the only J_2 that one can be coupled to J_1 to give $J = 1$ are $J_2 = J_1 \pm 1$, one concludes that $|0,(\gamma);1M\rangle = 0$ and consequently all $|0,(\gamma);JM\rangle$ vanish identically.

The proof that Eq. (2.33) is the general answer goes by induction on J. It is true for $J = J_0, J_0 + 2, J_0 + 4$. Equation (2.25) gives us $f^{J+2}_{J_1J_2}$ in terms of $f^J_{J_1 \pm 2J_2}$ which are known by the induction hypothesis. Using those values given (2.33) one sees that the result obtained for $f^{J+2}_{J_1J_2}$ checks with Eq. (2.33). To finish the proof one checks that the expression of $f^J_{J_1J_2}$ given by (2.33) satisfies Eqs. (2.26)–(2.29).

D. Coupling of $[0,\mu_1]$ and $[0,\mu_2]$ to the identity

Since in the identity representation one has that the matrix elements of all generators are zero, one has that in (2.22), J must be zero and one must have $J_1 = J_2$ which implies $\mu_1 = \mu_2$. For the $f^0_{J_1J_1}$ one obtains a set of equations which differs from the set (2.25)–(2.29) by having all the RHS equal to zero. This set of equations is easily solved giving

$$f^0_{JJ} = \lambda\,(-)^{(J-J_0)/2}[(2J+1)(2J-1)]^{1/2}, \qquad (2.34)$$

where λ is an arbitrary constant and J_0 is the minimum of J of irrep $[0,\mu]$.

3. CONCLUDING REMARKS

The most surprising result found in the explicit calculations of the last section is that the self-conjugate multiplicity-free irreps do not couple to form general multiplicity-free irreps. Stated more provocatively, no matter how many quarkels are coupled together (provided the number is odd and finite) the multiplicity-free part of the result still has the spectrum of a quarkel. Such a curious property must have a more instructive proof than our previous direct calculation. We have indeed found a more elegant proof which we now present.

First recall that under conjugation the third rank invariant $\mathscr{I}_3$ reverses sign ($\mathscr{C}:\mathscr{I}_3 \to -\mathscr{I}_3$), so that the self-conjugate irreps are characterized by $\mathscr{I}_3 = 0$.

Recall also that the third rank invariant is found by first coupling the generators to form the symmetric adjoint operator $\mathbf{X}$ s $\mathbf{X}$ and then forming an invariant with $\mathbf{X}$, i.e., $\mathscr{I}_3 = \mathbf{X}\cdot(\mathbf{X}$ s $\mathbf{X})$. [The symmetry of these products shows that $\mathscr{I}_3 = (\mathbf{X}$ s $\mathbf{X})\cdot\mathbf{X}$ also.]

The significant fact, which we will use in the demonstration to follow, is that for the discrete self-conjugate (multiplicity free) irreps not only does $\mathscr{I}_3 = 0$ but also *the symmetric adjoint operator* $\mathbf{X}$ s $\mathbf{X}$ *itself vanishes identically.*[23]

Consider now the coupling of two kinematically independent such systems, with generators $\mathbf{A}$ and $\mathbf{B}$ to form the composite generator $\mathbf{X} = \mathbf{A} + \mathbf{B}$, where it is assumed that $[\mathbf{A},\mathbf{B}] = 0$. The symmetric adjoint operator now takes the form $\mathbf{X}$ s $\mathbf{X} = 2\mathbf{A}$ s $\mathbf{B}$, since $\mathbf{A}$ s $\mathbf{A} = \mathbf{B}$ s $\mathbf{B} = 0$. Forming the invariant $\mathscr{I}_3$ leads to

$\mathscr{I}_3$ (composite system) $= 2(\mathbf{A} + \mathbf{B})\cdot(\mathbf{A}$ s $\mathbf{B})$.

Using the symmetry of the dot- and s -product we see that $\mathbf{A}\cdot(\mathbf{A}$ s $\mathbf{B}) = (\mathbf{A}$ s $\mathbf{A})\cdot\mathbf{B} = 0$ and similarly for the term in $\mathbf{B}$, $\mathbf{B}\cdot(\mathbf{A}$ s $\mathbf{B}) = \mathbf{B}\cdot(\mathbf{B}$ s $\mathbf{A}) = (\mathbf{B}$ s $\mathbf{B})\cdot\mathbf{A} = 0$.

We conclude that $\mathscr{I}_3$ (composite system) $= 0$, and hence that only self-conjugate irreps can be formed in the product.

The (multiplicity-free) coupling of two quarkels thus yields only the (bosonic) irreps $[0,(\beta)]$ and $[0,(\gamma)]$. Coupling in a *third* quarkel then yields only the irrep $[0,(\alpha)]$. Although this composite irrep is indistinguishable under the group from the quarkel itself, it should be noted that—by using a quarkel-counting $\overline{\mathrm{SL}(3,R)}$ invariant operator, for example—the two systems *can* be distinguished. (Moreover, the structure of the composite spin-$\frac{1}{2}$ state inherently involves states with spins greater than $\frac{1}{2}$.) Coupling in more quarkel systems leads to a repetition of the couplings already found.[24]

Only for a system having infinitely many states is such a curious property ("stability under many-particle-coupling") possible. (See note added in proof at end of paper.) It is intriguing to observe that such an unusual structure allows a model in which the existence of new and heavier "fundamental quarks" is unnecessary to hypothesize, such structures appearing quite naturally as composites of quarkels.

TABLE I. Expected values of μ_3 in Eq. (2.22) obtained by considering only the angular momentum content of $[0,\mu_1]$ and $[0,\mu_2]$.

μ_1	μ_2	μ_3	
(α)	(α)	(β)	1
		(γ)	2
(α)	(β)	(α)	3
(α)	(γ)	(α)	4
(β)	(β)	(β)	5
		(γ)	6
(γ)	(γ)	(β)	7
		(γ)	8
(γ)	(β)	(γ)	9

ACKNOWLEDGMENTS

One of us (JACA) wants to acknowledge financial support from the Fundacão dè Amparo a Pesquisa do Estado de São Paulo (FAPESP), São Paulo, Brazil, during his stay at Duke University where most of this work was done.

Note added in proof: Our argument is complete only under the restriction to multiplicity-free irreps in resolving the coupling. Since there exist both discrete and continuum self-conjugate irreps, which are not multiplicity free, the possibility is open that more general self-conjugate irreps might occur.

[1]Y. Dothan, M. Gell-mann, and Y. Ne'eman, Phys. Lett. **17**, 148 (1965).
[2]O.L. Weaver and L.C. Biedenharn, Nucl. Phys. A **185**, 1 (1972).
[3]O.L. Weaver, R.Y. Cusson, and L.C. Biedenharn, Ann. Phys. (N.Y.) **102**, 493 (1976).
[4]G. Rosensteel and D.J. Rowe, Ann. Phys. (N.Y.) **96**, 1 (1976).
[5]B. Buck, L.C. Biedenharn, and R.Y. Cusson, Nucl. Phys. A **317**, 205–41 (1979).
[6]R.F. Dashen and M. Gell-Mann, Phys. Lett. **17**, 145, 146 (1965).
[7]Y. Ne'eman, Ann. Inst. Henri Poincaré, A **28**, 396 (1978).
[8]F.W. Hehl, E.A. Lord, and Y. Ne'eman, Phys. Lett. B **71**, 432 (1977); Phys. Rev. D **17**, 428 (1978). These authors propose to call the quarkel irrep by the name "bandor."
[9]Dj. Sijacki, J. Math. Phys. **16**, 298 (1975).
[10]We propose the name "quarkel" for this spinorial irrep. This name is apt since (in German) "quark" designates a kind of cheese and there exists (in Hessen) a cheese (having a curious layered structure) called "quarkel." Professor Gell-Mann also suggested this name to us.
[11]This is not directly a result of the self-conjugate property. [An easy counterexample occurs in SU(3): the octet is self-conjugate, but the coupling of two-octets can yield a decouplet, which is not self-conjugate.]
[12]A. Bohr and B.R. Mottelson, in *Nuclear Structure* (Benjamin, Reading, Massachusetts, 1975), Vol. I.
[13]L.C. Biedenharn, R.Y. Cusson, M.Y. Han, and O.L. Weaver, Phys. Lett. B **42**, 257 (1972); Erratum, Phys. Lett. B **73**, 503 (1978).
[14]D.W. Joseph, "Representations of the Algebra of SL(3R) with $\Delta j = 2$," U. Nebraska preprint, February 1970 (unpublished).
[15]O.L. Weaver, Ph.D. Thesis, Duke University (1970).
[16]It is appropriate here to mention that there is no 3/2-band. This was first pointed out by Joseph[14] and (incorrectly) criticized in Ref. 13. An erratum correcting this error was submitted to Phys. Lett. late in 1973, but was inexplicably not published (though received by the editors) until recently [Phys. Lett. B **73**, 503 (1978)]. Several authors have since given the correct answer[25] but the error still appears occasionally in the literature.[26]
[17]H. van Dam and L.C. Biedenharn, Phys. Rev. D **14**, 405 (1976).
[18]L.C. Biedenharn, J.M. Blatt, and M.E. Rose, Rev. Mod. Phys. **24**, 249 (1952); D.M. Brink and G.R. Satchler, *Angular Momentum* (Clarendon, Oxford, 1971).
[19]The irrep α is clearly discrete. We designate the β and γ irreps as "discrete" since the invariants $\mathscr{I}_2 \to -\frac{1}{4}$ and $\mathscr{I}_3 \to 0$ constitute an isolated limit point in the continuum, separating the two continuum irreps having $\eta = \pm|\eta|$, $0 < |\eta| < \infty$.
[20]Other realizations in terms of particle and antiparticle operators are also possible.
[21]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1974).
[22]The standard notation $a! = a(a-1)(\cdots)$, $a!! = a(a-2)(\cdots)$ suggests the notational generalization: $a!^k$ for $a(a-k)(\cdots)$.
[23]The vector part of $\mathbf{X}$ s $\mathbf{X}$ must be a linear combination of the *symmetrized* bilinear products: $J \otimes J$, $J \otimes T$, and $T \otimes T$. The first and third of these, using the symmetry requirement, contain no vector part. The second of these product also contains no vector part, since T induces only $j \to j \pm 2$ transitions (and J is diagonal: $j \to j$); hence $J \otimes T \to (j \to j \pm 2)$, which is impossible for a vector operator. Hence $\mathbf{X} \otimes \mathbf{X}$ vanishes identically for that realization of the generators, $\mathbf{X}$, associated with η identically zero.
[24]Alternatively one could introduce a quasispin structure[27] which would couple inert pair states into the wave function. Since the "shell structure" is infinite, this would lead to an infinite quasispin.
[25]V.U. Ogievetzkii and E. Sokachev, Theor. Math. Phys. **22**, 462 (1975).
[26]Y. Güler, J. Math. Phys. **19**, 508 (1978).
[27]M.H. MacFarlane, "Shell Model Theory of Identical Nucleons," in *Lectures in Theoretical Physics* (U. Colorado Press, Boulder, Colorado, 1966); L.C. Biedenharn and J.D. Louck, in *Angular Momentum in Quantum Physics: Theory and Applications* (Addison–Wesley, Reading Massachusetts, 1979).

Comment on the Wigner 9-*j* symbol

R. Chatterjee and J. M. Dixon[a)]
Physics Department, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

(Received 21 December 1978)

A number of errors in the algebraic formulas of Rotenberg *et al*'s *The 3-j and 6-j Symbols* have been corrected. Attention is drawn to a few other simple new algebraic relationships.

The importance of the 9-*j* symbol is well known from the work of Wigner quoted by Jahn and Hope[2] and is used extensively in the atomic theory of angular momentum by Yutsis *et al.*[3] and many others.[4] It appears in the calculation of the matrix element of any double tensor when the Wigner-Eckart theorem is applied. A number of tables exist such as those of Jucys,[5] Howell,[6] and Landolt-Börnstein.[7] Stassis and Williams[8] have already corrected one of the algebraic formulae, involving a 9-*j* symbol, quoted in Rotenberg *et al.*[9] In this note we will point out other errors quoted by Rotenberg *et al*,[9] correct them, and provide a few new simple relationships.

The 9-*j* symbol (square brackets) can be expressed either in terms of 6-*j* symbols (curly brackets) or 3-*j* symbols (round brackets). Thus

$$\begin{bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \\ j_7 & j_8 & j_9 \end{bmatrix} = \sum_j (-1)^{2j}(2j+1) \begin{Bmatrix} j_1 & j_4 & j_7 \\ j_8 & j_9 & j \end{Bmatrix} \begin{Bmatrix} j_2 & j_5 & j_8 \\ j_4 & j & j_6 \end{Bmatrix} \begin{Bmatrix} j_3 & j_6 & j_9 \\ j & j_1 & j_2 \end{Bmatrix} \tag{1}$$

$$= \sum_{\text{all } m} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} j_4 & j_5 & j_6 \\ m_4 & m_5 & m_6 \end{pmatrix} \begin{pmatrix} j_7 & j_8 & j_9 \\ m_7 & m_8 & m_9 \end{pmatrix} \begin{pmatrix} j_1 & j_4 & j_7 \\ m_1 & m_4 & m_7 \end{pmatrix} \begin{pmatrix} j_2 & j_5 & j_8 \\ m_2 & m_5 & m_8 \end{pmatrix} \times \begin{pmatrix} j_3 & j_6 & j_9 \\ m_3 & m_6 & m_9 \end{pmatrix}. \tag{2}$$

There are two major applications of the 9-*j* symbol in physics. The first is to transform a state in a *j*–*j* coupling scheme to an L–S coupling, which is given by

$$|(s_1\ l_1)\ j_1\ (s_2\ l_2)\ j_2, JM\rangle = \sum_{SL} [(2S+1)(2L+1)(2j_1+1)(2j_2+1)]^{1/2} \begin{bmatrix} s_1 & s_2 & S \\ l_1 & l_2 & L \\ j_1 & j_2 & J \end{bmatrix} \times |(s_1\ s_2)\ S\ (l_1\ l_2)\ L, JM\rangle.$$

The second important application of the 9-*j* symbols is to evaluate the reduced matrix elements of the double tensor $W^{(k_1k_2)k}$ defined by Judd.[10] The latter many-electron double tensor is defined with rank k_1 in spin space and rank k_2 in orbital space, k being the total rank of the tensor. The reduced matrix elements are given by

$$\langle \gamma\ s_1\ l_1\ j_1 || W^{(k_1k_2)k} || \gamma'\ s_2\ l_2\ j_2\rangle = [(2j_1+1)(2j_2+1)(2k+1)]^{1/2} \times \begin{bmatrix} s_1 & l_1 & j_1 \\ s_2 & l_2 & j_2 \\ k_1 & k_2 & k \end{bmatrix} \langle \gamma\ s_1\ l_1 || W^{(k_1k_2)} || \gamma'\ s_2\ l_2\rangle \tag{3}$$

The reduced matrix elements $\langle \gamma\ s_1\ l_1 || W^{(k_1k_2)} || \gamma'\ s_2\ l_2\rangle$ has been tabulated by Nielson and Koster[11] with $k_1 = 1$, $k_2 = 1$ for the p^n, d^n, and f^n configurations. Chatterjee *et al.*[12] have extended this tabulation for $k_1 = 1$ and $k_2 = 0$ to 6 for all these configurations.

We have re-examined the 9-*j* symbol using Eq. (2) in the light of Regge symmetry. We have found that Rotenberg (R) *et al.*'s Eqs. 1.11 and 1.12 are incorrect. The correct forms, obtained from Regge symmetry, are given below:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_1+j_2+j_3} \begin{pmatrix} \tfrac{1}{2}(j_2+j_3+m_1) & \tfrac{1}{2}(j_1+j_3+m_2) & \tfrac{1}{2}(j_1+j_2+m_3) \\ j_1-\tfrac{1}{2}(j_2+j_3-m_1) & j_2-\tfrac{1}{2}(j_1+j_3-m_2) & j_3-\tfrac{1}{2}(j_1+j_2-m_3) \end{pmatrix}$$

replaces Eq. 1.11 (R) and

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} j_1 & \tfrac{1}{2}(j_2+j_3+m_1) & \tfrac{1}{2}(j_2+j_3-m_1) \\ j_2-j_3 & -m_3-\tfrac{1}{2}(j_2-j_3+m_1) & +m_3-\tfrac{1}{2}(j_2-j_3-m_1) \end{pmatrix}$$

replaces equation 1.12 (R). To derive these relations, we have used the connection between a 3-*j* symbol and Regge's matrix,

[a)]Visiting Associate Professor at the University of Calgary during 1978/79.

 0022-2488/79/081622-02$01.00

then used the reflection symmetry about the major diagonal of this matrix, and converted from the new Regge matrix to a 3-j symbol by solving nine equations.

Rotenberg *et al.*'s[1] Eq. (3.8) should be replaced by the following correct equation given by Kennedy *et al.*[13]:

$$\sum_{\substack{d,e,f \\ \text{all even}}} (2d+1)(2e+1)(2f+1)\begin{bmatrix} A & B & C \\ A & B & C \\ d & e & f \end{bmatrix}^2 = \frac{1}{4}\left[1 + \frac{(-1)^{2A}}{2A+1} + \frac{(-1)^{2B}}{2B+1} + \frac{(-1)^{2C}}{2C+1}\right]. \tag{4}$$

Landolt-Börnstein[7] also give Rotenberg *et al.*'s[1] erroneous form for this latter equation.

Rotenberg *et al.*'s[1] Eq. 3.20 should be corrected by the following equation:

$$\begin{bmatrix} 0 & b & c \\ d & e & f \\ d & e & f \end{bmatrix} = (-1)^{b+d+e+f}[(2d+1)(2b+1)]^{-1/2}\begin{Bmatrix} b & f & f \\ d & e & e \end{Bmatrix}\delta(b,c), \tag{5}$$

which is easily obtained from Eq. 3.19(R).

In Rotenberg *et al.*'s[1] Eq. 3.19, if we put $f = 0$, then we obtain the following relationship:

$$\begin{bmatrix} a & b & e \\ a & b & e \\ 0 & 0 & 0 \end{bmatrix} = (-1)^{2(a+b+e)}[(2a+1)(2b+1)(2e+1)]^{-1/2}. \tag{6}$$

This expression (6) will be useful to calculate the reduced matrix elements of the double tensor $W^{(00)0}$. In Rotenberg's Eq. 3.19, if we put $d = 0$, we get the following equation:

$$\begin{bmatrix} a & b & e \\ e & 0 & e \\ b & b & 0 \end{bmatrix} = (-1)^{a-b-e}[(2b+1)(2e+1)]^{-1}. \tag{7}$$

From Rotenberg *et al.*'s[1] Eq. 3.22, if we put $k = 0$, we get

$$\sum_x (2x+1)(-1)^x \begin{bmatrix} a & b & x \\ c & d & e \\ f & g & h \end{bmatrix} = (-1)^{d+f-2c}[2c+1]^{-1}\delta(b,e)\delta(c,g)\delta(a,h). \tag{8}$$

ACKNOWLEDGMENTS

One of the authors (J. M. D.) would like to express his thanks to the Physics Department of the University of Calgary for support and the members of this department for their kindness and hospitality during his stay. He would also express his appreciation to Professor H.A. Buckmaster for extending the facilities of the EPR group to him. This work is part of a research program supported by the National Research Council of Canada and the University of Calgary.

[1]E.P. Wigner, (1940) (unpublished), quoted in *Quantum Theory of Angular Momentum* edited by L.C. Biedenharn and H. Van dam (Academic, New York, 1965).
[2]H.A. Jahn, and J. Hope, Phys. Rev. **93**, 318 (1954).
[3]A.P. Yutsis, *et al.*, "Theory of Angular Momentum", Israel Program for Scientific Translations, Jerusalem (1962).
[4]J.P. Elliott, and A.M. Lane, Encycl. Phys. **39**, 241 (1957).
[5]A. Jucys, *Table of 9j coefficients* (Computing Center AS USSR, Moscow, 1968).
[6]K.M. Howell, "Tables of 9j symbols," University of Southampton Research Dept. 59-2 (1959).
[7]Landolt-Börnstein, in H. Apel, *Numerical Tables for Angular Correlation Computations in α-, β- and γ-Spectroscopy; 3j, 6j and 9j Symbols F- and Γ-Coefficients*, edited by H. Schopper (Springer-Verlag, Berlin, 1968), Vol. 3.
[8]C. Stassis and S.A. Williams, J. Math. Phys. **17**, 480 (1976).
[9]M. Rotenberg, R. Bivins, N. Metropolis and J.K. Wooten, Jr., *The 3-j and 6-j Symbols* (The Technology Press, MIT, Cambridge, Mass., 1959).
[10]B.R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[11]C.W. Nielson and G.F. Koster, *Spectroscopic Coefficients for the p", d" and f" Configurations* (MIT press, Cambridge, Mass., 1963).
[12]R. Chatterjee, M.R. Smith, and H.A. Buckmaster, Can. J. Phys. **54** (12), 1228–33 (1976).
[13]J.M. Kennedy, B.J. Sears, and W.T. Sharp, "Tables of X Coefficients," CRT-569, Atomic Energy of Canada Ltd., No. 106 (1954).

Representation matrix elements and Clebsch–Gordan coefficients of the semisimple Lie groups

A. U. Klimyk and A. M. Gavrilik

Institute for Theoretical Physics, Academy of Sciences of the Ukrainian SSR, Kiev, USSR
(Received 21 January 1976)

We give a general theory of matrix elements (ME's) of the unitary irreducible representations (UIR's) of linear semisimple Lie groups and of reductive Lie groups. This theory connects together the following things, (1) MEUIR's of all the representation series of a noncompact Lie group, (2) MEUIR's of compact and noncompact forms of the same complex Lie group. The theory presented is based on the results of the theory of the principal nonunitary series representations and on a theorem which states that ME's of the principal nonunitary series representations are entire analytic functions of continuous representation parameters. The principle of analytic continuation of Clebsch–Gordan coefficients (CGC's) of finite dimensional representations to CGC's of the tensor product of a finite and an infinite dimensional representation and to CGC's of the tensor product of two infinite dimensional representations is proved. ME's for any UIR of the group $U(n)$ and of the group $U(n,1)$ are obtained. The explicit expression for all CGC's summed over the multiplicity of the irreducible representation in the tensor product decomposition is derived.

I. INTRODUCTION

The representations of some semisimple and reductive Lie groups have become of increased importance in physics. The compact groups $U(n)$ and $SO(n)$ are extensively used in the elementary particle physics[1] and in the nuclear spectroscopy.[2] Among the applications of the $SO_0(n,1)$ representations there are the dynamical properties of the hydrogen atom[3] and the quantum field theory.[4] The $U(n,1)$ representations are applied to the N-dimensional harmonic oscillator and to the hadron spectroscopy.[5] Fom the point of view of physicists he most important aspects of the representation theory are MEUIRs and CGCs. MEUIRs and CGCs for compact and noncompact Lie groups, as well as for different representation series (principal unitary, supplementary unitary, discrete) of the same noncompact group, were usually studied separately. But there exist close relations among them. Such relations were searched after the explicit form of MEUIRs and CGCs had been found. There are two main results in the paper. Firstly, the general theory of MEs of irreducible representations (unitary and nonunitary) for semisimple Lie groups is given. This theory connects (1) MEUIRs of all the representation series of a noncompact semisimple Lie group, (2) MEUIRs of compact and noncompact forms of the same complex semisimple Lie group. According to this theory one does not need to derive MEUIRs of each representation series separately. This theory gives unified approach to MEUIRs of all the series. Moreover, MEUIRs of any series may be easily obtained provided MEUIRs of some continuous series are known. We show how this theory does work considering the groups $U(n,1)$, $U(n)$, $SO_0(n,1)$, $SO(n)$. Our approach is based on the theory of the principal nonunitary series representations.

The principal nonunitary series representations will be called the elementary representations.[6] This series contains the principal unitary series representations. Moreover, the elmentary representations are obtained by "analytic continuation" of the principal unitary series.

The elementary representations of semisimple Lie groups were studied in the papers by Harish-Chandra. The list of these papers as well as their account can be found in Ref. 7. It was proved that any completely (or infinitesimally) irreducible representation of a linear semisimple Lie group is a subquotient of some elementary representation. In particular, the elementary representations contain finite dimensional irreducible representations. Therefore, we obtain MEs of any irreducible representation as MEs of some elementary representation.

This paper proves the theorem which states that MEs of elementary representations are entire analytic functions of continuous representation parameters. This theorem allows one to obtain MEs of any representation series by means of analytic continuation of MEs of some nondegenerate representation series.

Most of unitarizable irreducible representations enter the elementary representations not being unitarized. Therefore, to find MEUIRs from MEs of the elementary representations, one has to change the basis. The explicit form of the infinitesimal operators in the initial basis and in the new one can be used to obtain the transition matrix.

MEUIRs possess numerous symmetries which are of great importance for applications. Some symmetries are described by means of the so called intertwining operators of elementary representations. These symmetries lead to some relations for the special functions. There exists a one-to-one correspondence between the intertwining operators and the Weyl group elements. That is why the Weyl group as well as the correspondence mentioned are considered.

The theory of MEUIRs and MEs of the elementary representations allows one to apply the procedure of analytic continuation to study CGCs of semisimple Lie groups. It gives a possibility to continue analytically CGCs of finite dimensional representations to CGCs of the tensor product of a finite and an infinite dimensional representations or of the tensor product of infinite dimensional representations.

 0022-2488/79/081624-19$01.00

The tensor product of a finite and an infinite dimensional representations is of increased interest both in mathematics and physics.[8] The Wigner–Eckart theorem for noncompact semisimple groups[9] deals with this tensor product. Segal considers that a finite mass spectrum can be obtained by decomposing the tensor product of a finite and an infinite dimensional representations of the group $SO_0(3,2)$ (cf. Introduction in Ref. 8).

The information on the tensor product of infinite dimensional representations [except for the groups SO(2,1) and SO(3,1)] is rather poor.[10] In particular, the problem of the multiplicity of irreducible representation in this tensor product is not yet solved. Therefore, we don't know whether analytic continuation of CGCs of finite dimensional representations gives all CGCs for the tensor product of infinite dimensional representations. We hope that the procedure of analytic continuation of CGCs of finite dimensional representations will make it possible to obtain an estimate for the multiplicity of irreducible representation in the tensor product of infinite dimensional representations.

Our theory of MEUIRs and CGCs uses the theory of the principal nonunitary series representations. These representations are induced by irreducible representations of the minimal parabolic subgroup (for the definitions see Ref. 7). Irreducible representations of other parabolic subgroups induce the degenerate nonunitary series representations. The theory of MEs and CGCs of the corresponding degenerate unitary series representations can be constructed in the same way.

The second main result in the paper is the derivation of all MEs of all UIRs of the groups $U(n,1)$ and $U(n)$. In Ref. 11 the integral form of MEUIRs of $U(n)$ and of MEs of the principal unitary series of $U(n,1)$ was obtained. We obtain the explicit form of MEUIRs of $U(n)$ and of $U(n,1)$ with the help of our general theory.

MEUIRs of $U(n)$ allow us to obtain CGCs of this group summed over the multiplicity of the irreducible representation in the tensor product decomposition. If this multiplicity is 1 we have the explicit expression of CGCs. In this way one obtains some CGCs of $U(n)$ not yet known.[12]

We also give the relations between MEUIRs of different representation series of $SO_0(n,1)$.

II. REPRESENTATIONS OF THE PRINCIPAL NONUNITARY SERIES AND THEIR MATRIX ELEMENTS

In this section we give necessary notions from the theory of semisimple (and reductive) Lie groups and prove that MEs of the principal nonunitary series representations are entire functions.

Let G be a connected linear semisimple or reductive Lie group, and $\mathfrak{g}$ its Lie algebra. Let K be a maximal compact subgroup of G, and $\mathfrak{j}$ a subalgebra of $\mathfrak{g}$ which corresponds to K. If $B(\cdot,\cdot)$ is the Killing–Cartan form on $\mathfrak{g}$, then let $\mathfrak{p}$ be the orthogonal completion to $\mathfrak{j}$ in $\mathfrak{g}$ with respect to $B(\cdot,\cdot)$. Then $\mathfrak{g} = \mathfrak{j} + \mathfrak{p}$. If θ is the Cartan involution on $\mathfrak{g}$, then the form $(x,y) = -B(x,\theta y)$, is a positive definite scalar product on $\mathfrak{g}$. Let a_+ be a maximal commutative subalgebra in $\mathfrak{p}$. The dimension of a_+ is called the split rank of $\mathfrak{g}$. We denote by $\mathfrak{m}$ the centralizer of a_+ in K. If a_- is a Cartan subalgebra of $\mathfrak{m}$, then $a = a_+ + a_-$ is a Cartan subalgebra of $\mathfrak{g}$. Consider a set of operators $\mathrm{ad}H$, $H \in a_+$, which act in the space $\mathfrak{g}$. If $\mathfrak{g}$ is equipped with a scalar product (x,y), then the operators $\mathrm{ad}H$ are a commutative family of self-adjoint operators. Therefore, $\mathfrak{g}$ is decomposed into a direct sum of eigensubspaces of $\mathrm{ad}H$, $\mathfrak{g} = \mathfrak{g}_0 + \Sigma_\lambda \mathfrak{g}_\lambda$. The sum is over nonzero linear forms on a_+, and $\mathfrak{g}_0$ corresponds to the eigenvalue 0. The forms λ are called the restricted roots of $(\mathfrak{g},a_+)$. The roots λ are divided into positive ($\lambda > 0$) and negative ($\lambda < 0$) ones. If $\mathfrak{n} = \Sigma_{\lambda>0}\mathfrak{g}_\lambda$, then $\mathfrak{n}$ is a maximal nilpotent subalgebra in $\mathfrak{g}$. Let N, A_+ be the analytic subgroups in G with the Lie algebras $\mathfrak{n}$, a_+, respectively. Then the Iwasawa decomposition $G = A_+NK$ is valid, and each element g, $g \in G$, is uniquely decomposed into a product $g = h_+nk$, $h_+ \in A_+$, $n \in N$, $k \in K$. The map $(h_+,n,k) \to h_+nk$ is an analytic diffeomorphism from the manifold $A_+ \times N \times K$ onto G.

Let M be the centralizer of A_+ in K and M^* the normalizer of A_+ in K. Then the quotient M^*/M is a finite group which is called the Weyl group W of the pair $(\mathfrak{g},a_+)$. The action of the elements of W upon linear forms Λ, $\Lambda \in (a_+)^*$, and upon finite dimensional representations δ of M is defined. Namely, if $w \in W$ is a coset m^*M, $m^* \in M^*$, then

$$(w\Lambda)(H) = \Lambda(m^{*-1}Hm^*), \quad H \in a_+, \tag{1}$$

$$(w\delta)(m) = \delta(m^{*-1}mm^*), \quad m \in M. \tag{2}$$

Moreover, $w\Lambda$ and $w\delta$ do not depend on the choice of the representative m^* in m^*M.

Let δ be a finite dimensional representation of M in the space H_δ, and let Λ be a complex linear form on a_+. Next, let $L^2_\delta(K,H_\delta)$ be a Hilbert space of all measurable functions f: $K \to H_\delta$ such that

$$\int_K \|f(k)\|^2_{H_\delta} dk < \infty, \quad f(mk) = \delta(m)f(k), \tag{3}$$

where $m \in M$, $k \in K$, and dK is the invariant measure on K. Then the operators $\pi_{\delta,\Lambda}(g)$,

$$\pi_{\delta,\Lambda}(g)f(k) = \exp[\Lambda(\log h_+)]f(k_g),$$

where $kg = h_+nk_g$, $h_+ \in A_+$, $n \in N$, and $k_g \in K$, define a representation of G. The set of all representations $\pi_{\delta,\Lambda}(g)$ is called the principal nonunitary series. This series is the analytic continuation of the principal unitary series representations. Thus, the principal unitary series is contained in the principal nonunitary series. The representation $\pi_{\delta,\Lambda}$ will be called the elementary representation. It is shown in Ref. 13 that every completely (or infinitesimally) irreducible representation of G which is decomposable into a direct sum of unitary irreducible representations of K with a finite multiplicity is infinitesimally equivalent to a subquotient of some elementary representation. For the definitions of the infinitesimal equivalence and the infinitesimal irreducibility see Ref. 7.

Since the action of the elements w of W upon δ and Λ is defined the action of w in the set of elementary representations is also defined, namely, $w\pi_{\delta,\Lambda} = \pi_{w\delta,w(\Lambda-\rho)+\rho}$.

The condition (3) shows that the space $L^2_\delta(K,H_\delta)$ may be identified with the subspace of the space $L^2(K)$. The space $L^2_\delta(K,H_\delta)$ is decomposed into orthogonal sum of subspaces in which $\pi_{\delta,\Lambda}$ realizes irreducible representations of K. The direct sum of these subspaces is denoted by $dL^2_\delta(K,H_\delta)$. The representation $\pi_{\delta,\Lambda}$ of $\mathfrak{g}$ in $dL^2_\delta(K,H_\delta)$ will be denoted by $d\pi_{\delta,\Lambda}$.

Every function from $dL^2_\delta(K,H_\delta)$ is infinitely differentiable. This is a consequence of the analyticity of functions of $dL^2_\delta(K,H_\delta)$ with respect to $\pi_{\delta,\Lambda}$ (see Corollary 4.4. 5.17 in Ref. 7).

Let l be the split rank of G. The form Λ on a_+ is defined by l numbers if some basis is chosen in a_+. We denote these numbers by $c_1,c_2,...,c_l$. Thus $\pi_{\delta,\Lambda}\equiv\pi_{\delta,c_1,c_2,...,c_l}$. Let $\mathfrak{M}_\delta$ be the set of the representations $\pi_{\delta,\Lambda}$ of G with fixed δ.

Theorem 1: Matrix elements of the representations $\pi_{\delta,c_1,c_2,...,c_l}$ from $\mathfrak{M}_\delta$ in a fixed basis of $dL^2_\delta(K,H_\delta)$ are entire analytic functions of the parameters $c_1,c_2,...,c_l$.

Proof: Let $f_1(k)$ and $f_2(k)$ be two basis elements in $dL^2_\delta(K,H_\delta)$. Consider ME

$$\int (f_1(k),\exp[\Lambda(\log h_+)]\, f_2(k_g))\, dk$$

$$= \int (f_1(k),\pi_{\delta,c_1,c_2,...,c_l}(g) f_2(k))\, dk,$$

where g is a fixed element of G, and $(\cdot,\cdot)$ is a scalar product in H_δ, The function $(f_1(k),\exp[\Lambda(\log h_+)]\, f_2(k_g))$ is an entire analytic function of the parameters $c_1,c_2,...,c_l$. From the differentiability of the integral in a parameter which the integral depends on, from the infinite differentiability of functions of $dL^2_\delta(K,H_\delta)$ and from the differentiability of the representation $\pi_{\delta,c_1,c_2,...,c_l}$ (for the definition of the latter see Ref. 7) it follows that the ME is an entire function.

III. REPRESENTATIONS OF THE PRINCIPAL NONUNITARY SERIES AND FINITE DIMENSIONAL REPRESENTATIONS

It was mentioned in Sec. II that every completely irreducible representation of G is infinitesimally equivalent to a subquotient of some elementary representation. Hence, every finite dimensional representation of G is contained in some elementary representation. Let λ be a weight of the finite dimensional representation ω of G with respect to the Cartan subalgebra a. Let $\lambda(\omega)$ be a restriction of the weight λ onto a_+. $\lambda(\omega)$ will be called the restricted weight. Let $\Lambda(\omega)$ denote the lowest restricted weight of ω. Let ω_δ denote a finite dimensional representation of G which is an extension of the fixed representation δ of M. In other words ω_δ is a finite dimensional representation of G for which the space of all weight vectors belonging to all the weights λ with $\lambda(\omega_\delta)=\Lambda(\omega_\delta)$, is invariant with respect to M and is a space of the representation δ. The following theorem is proved in Ref. 14.

Theorem 2: The elementary representation $\pi_{\delta,\Lambda}$ of G may contain only one finite dimensional subrepresentation, this being with a multiplicity not exceeding one. Moreover, $\pi_{\delta,\Lambda}$ may contain as a finite dimensional subrepresentation only the representation ω_δ (an extension of the representation δ of M). The representation $\pi_{\delta,\Lambda}$ contains ω_δ as a subrepresentation if and only if $\Lambda(\omega_\delta)=\Lambda$.

Let δ be a fixed unitary irreducible representation of M, and let N be a set of positive integers. A set of $\Lambda\in a^*_+$ (a^*_+ is a dual of a_+) such that $\Lambda=\Lambda(\omega_\delta)$ is a lattice in a^*_+ which contains an infinite numbers of points. This lattice is in a one-to-one correspondence with the lattice N^l (l is the split rank of G) from which a finite set of points is thrown out.

Theorems 1 and 2 imply the following corollary.

Corollary: Matrix elements of a finite dimensional representation of G in an appropriate basis are the matrix elements of a corresponding elementary representation of G (this correspondence is given by Theorem 2). Matrix elements of elementary representations of G are an analytic continuation of matrix elements of its finite dimensional representations in an appropriate basis.

This analytic continuation is not unique. Therefore, when using the analytic continuation one must be certain that it is the necessary one. The guarantee is achieved by means of some additional condition. Fulfillment of the relations

$$\frac{d}{dt}g(t)=\mathrm{I}g(t) \tag{4}$$

for noncompact one-parameter subgroups $g(t)$ with infinitesimal generators I serves as a sufficient condition for analytic continuation to be correct. For MEs the relations (4) are of the form

$$\frac{d}{dt}\langle n|g(t)|m\rangle=\sum_{n'}\langle n|\mathrm{I}|n'\rangle\langle n'|g(t)|m\rangle. \tag{5}$$

Hence, the relations (5) demand knowledge of the MEs of the operators I. If the basis elements in (5) belong to $dL^2_\delta(K,H_\delta)$, then the sum in (5) is finite (see Sec. XI below).

According to Theorem 1 MEs of, for example, the principle unitary series representations, define MEs of all elementary representations. Since elementary representations contain all completely irreducible representations, MEs of the principal unitary series define MEs of all other series of representations (in particular, of finite dimensional representations).

IV. MATRIX ELEMENTS AND INTERTWINING OPERATORS

MEs of elementary representations in an orthogonal basis of $L^2_\delta(K,H_\delta)$ give MEs of the representations of the principal unitary series in unitarized form. This is not true for the representations of the complementary and the discrete series because these are not unitary in the space $L^2_\delta(K,H_\delta)$ with defined scalar product. To obtain unitarized MEs it is necessary to introduce a new scalar product and to choose an orthonormal basis in a new Hilbert space. A transition to new scalar product can be realized by means of a self-adjoint operator A. Such operators are related to intertwining operators of elementary representations (see Lemma 22 and Pro-

position 25 in Ref. 15). If the explicit form of the operator A is available, it is possible to transform MEs of representations of the complementary, discrete, or other unitary series in orthonormal basis of the space $L^2_\delta(K,H_\delta)$ into MEs which are unitarized.

Besides that, intertwining operators define the symmetry relations for MEs of elementary and unitary representations as functions of the complex parameters $c_1,c_2,...,c_l$ of representations. These relations lead to some relations for special functions associated with the representations under consideration.

The explicit integral form for intertwining operators is introduced in Ref. 16. To apply the intertwining operators to MEs of representations it is necessary to have the intertwining operators in matrix form. Therefore, we give the following definition of intertwining operator. The linear operator $\Pi = \Pi^{\delta,\Lambda}_{\delta',\Lambda'}$ from $dL^2_\delta(K,H_\delta)$ into $dL^2_{\delta'}(K,H_{\delta'})$ such that

$$\Pi d\pi_{\delta,\Lambda} = d\pi_{\delta',\Lambda'}\Pi \tag{6}$$

is called the intertwining operator for the representations $\pi_{\delta,\Lambda}$ and $\pi_{\delta',\Lambda'}$.

The representations $\pi_{\delta,\Lambda}$ and $\pi_{\delta',\Lambda'}$ can possess a nonzero intertwining operator if they have common irreducible representations. The latter holds if $\pi_{\delta,\Lambda}$ and $\pi_{\delta',\Lambda'}$ have the same infinitesimal character.[17] In turn, it is possible iff the representation $\pi_{\delta',\Lambda'}$ is obtained from the representation $\pi_{\delta,\Lambda}$ by the action of some element of the Weyl group W_c of the complexification $[G]^c$ of G. Intertwining operators which correspond to elements of the Weyl group W of $(\mathfrak{g},a_*)$ are analytic functions[15] of the complex parameters $c_1,c_2,...,c_l$. The explicit matrix form of all the intertwining operators is found for all elementary representations[18-20] of the groups $U(n,1)$ and $SO_0(n,1)$ and will be given below.

According to the corollary in Sec. III MEs of the elementary representations lead to MEs of finite dimensional representations and, vice versa, MEs of finite dimensional representations can be analytically continued to MEs of elementary representations. For this reason sometimes it is necessary to have MEs of finite dimensional representations of G in an orthonormal basis of the space $L^2_\delta(K,H_\delta)$ if MEs of unitary finite dimensional representations of the compact form G_k of the complexification $[G]^c$ of G is known, and vice versa. The problem is to find explicitly the relation between these MEs. As it is known there is a one-to-one correspondence between the finite dimensional representations of G and G_k. This correspondence is realized as follows. MEs of finite dimensional representations of G and G_k are real analytic functions of group parameters. Parameters of the group G_k are continued analytically in the set of parameters of the group $[G]^c$ to parameters of the group G. If such continuation is done for MEs of a finite dimensional representation of G_k, then we obtain the corresponding MEs of a finite dimensional representation of G. This correspondence is reversible.

Suppose we have MEs of a finite dimensional representation of G in an orthonormal basis of $L^2_\delta(K,H_\delta)$. Continue these MEs in group parameters to MEs of a representation of G_k. The obtained representation matrices are not unitary. Unitarization of these matrices is achieved by a transition to a new basis. This transition is given by the operator A which is related to intertwining operator in the same way as in the case of the operator A which unitarizes the unitarizable representation of G. The explicit form of the operator A allows us to obtain unitarized MEs of finite dimensional representations of G_k and, vice versa, to transfer a basis in which the representation of G_k is unitary, into an orthonormal basis of the subspace of finite dimensional representation of the space $L^2_\delta(K,H_\delta)$.

The operator A can be found as follows. Either it can be calculated with the help of Lemma 22 and Proposition 25 of Ref. 15 [to do this one has to know the explicit form of the operator (9.6) in Ref. 15], or the operator A can be found by means of the fact that it realizes a unitarization of the representation both in global and infinitesimal form. In some cases all MEs of infinitesimal operators of elementary and unitary representations can be calculated. They allow one to write down a system of equations for MEs of the operator A. If this system is solved we can apply the operator A to global representations of G. In such way the operators A for the groups $U(n,1)$ and $SO_0(n,1)$ have been found explicitly. They are given below.

V. REPRESENTATIONS OF THE GROUP U(n,1)

We intend to apply the theory of MEs presented above to representations of the group $U(n,1)$. For this purpose here we shall describe elementary and irreducible representations of $U(n,1)$.

For $O(n,1)$, $K \approx U(n) \otimes U(1)$, $M \approx U(n-1) \otimes U(1)$, $A_* = \exp a_*$, where a_* is a one-dimensional subalgebra of the Lie algebra $u(n,1)$ [Lie algebra of $U(n,1)$] with the basis element $e = E_{n+1,n} + E_{n,n+1}$. Here and in what follows E_{ij} is an $(n+1)\times(n+1)$ matrix with elements $(E_{ij})_{ks} = \delta_{ik}\delta_{js}$. Thus the elementary representation of $U(n,1)$ is defined by the unitary irreducible representation δ of $U(n-1)\otimes U(1)$ and the linear form Λ on a_*. The representation δ is fixed by $n-1$ integers $m_1,m_2,...,m_{n-1}$ such that $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_{n-1}$ and by the integer m_0 which defines a character of $U(1)$. We shall also use the integers $l_1,l_2,...,l_{n-1}$ where $l_i = m_i - i - 1$. The linear form Λ is given by the number $\lambda = \Lambda(e)$. We shall use the numbers $c_1 = -\frac{1}{2}(\lambda - m_0 + 2)$ and $c_2 = \frac{1}{2}(\lambda + m_0 - 2n - 2)$ instead of the numbers m_0 and λ. Hence, the elementary representation is defined by the numbers $l_1,l_2,...,l_{n-1},c_1,c_2$. Therefore, we shall denote $\pi_{\delta,\Lambda}$ by $\pi(l_1,l_2,...,l_{n-1};c_1,c_2) \equiv \pi(l;c_1,c_2)$.

There exists only one nonidentical element w of the Weyl group W of the pair $(U(n,1),a_*)$, and $w\pi(l:c_1,c_2) = \pi(l:c_2,c_1)$.

The representation $\pi(l:c_1,c_2)$ of $U(n,1)$ decomposes into those and only those irreducible representations (with unit multiplicity) of $U(n)$, whose labels $m_{1n},m_{2n},...,m_{nn}$ $(m_{1n} \geqslant m_{2n} \geqslant \cdots \geqslant m_{nn})$ satisfy the condition

$$m_{1n} \geqslant m_1 \geqslant m_{2n} \geqslant m_2 \geqslant \cdots \geqslant m_{n-1,n} \geqslant m_{n-1} \geqslant m_{nn}. \tag{7}$$

Using this fact we choose a basis in the space of the represen-

tation $\pi(l;c_1,c_2)$ in such a way that it coincides with the union of orthonormal bases of subspaces of irreducible representations of the subgroup $U(n)$. Bases of these subspaces are chosen to be the Gel'fand–Zetlin bases, i.e., the bases in which the infinitesimal operators of $U(n)$ act according to the well-known Gel'fand–Zetlin formulas. We shall call this basis of the space of the representation $\pi(l;c_1,c_2)$ the canonical basis. The elements of the canonical basis will be labelled by the well-known Gel'fand–Zetlin schemes α and will be written as $|m_n,\alpha\rangle$, where $m_n=(m_{1n},m_{2n},...,m_{nn})$, and α consists of rows of integers labelling the irreducible representations of the subgroups $U(n-1)$, $U(n-2)$,...,$U(1)$. It is shown in Ref. 18 that the canonical basis can be chosen in such a way that the infinitesimal operators $E_{n,n+1},E_{n+1,n},E_{n+1,n+1}$ of the representation $\pi(l;c_1,c_2)$ are given by the formulas

$$E_{n,n+1}|m_n,\alpha\rangle = \sum_{s=1}^{n}(l_{sn}-c_1)\omega_s(l,m_n,\alpha)|m_n^{+s},\alpha\rangle, \tag{8}$$

$$E_{n+1,n}|m_n,\alpha\rangle = -\sum_{s=1}^{n}(l_{sn}-c_2-1)\omega_s(l,m_n^{-s},\alpha)|m_n^{-s},\alpha\rangle, \tag{9}$$

$$E_{n+1,n+1}|m_n,\alpha\rangle = (c_1+c_2+\sum_{i=1}^{n-1}l_i-\sum_{j=1}^{n}l_{jn}+n+1)|m_n,\alpha\rangle, \tag{10}$$

where

$$\omega_s(l,m_n,\alpha)=\Big|\prod_{j=1}^{n-1}(l_{j,n-1}-l_{sn}-1)^{1/2}(l_{sn}-l_j)^{1/2} \times\prod_{\substack{r=1\\r\neq s}}^{n}[(l_{sn}-l_{rn}+1)(l_{sn}-l_{rn})]^{-1/2}\Big|, \tag{11}$$

$l_{ij}=m_{ij}-i$, and $m_n^{\pm s}$ means a set of the numbers m_n, where $m_{sn}\pm 1$ is substituted instead of m_{sn}.

The representation $\pi(l;c_1,c_2)$ is irreducible if and only if c_1 and c_2 are not integers or if c_1 and c_2 coincide with some of the numbers $l_1,l_2,...,l_{n-1}$. The representations $\pi(l;c_1,c_2)$ for which $c_1=\bar{c}_2$ and only those belong to the principal unitary series. The structure (composition series) of reducible elementary representations is given in Ref. 21. In particular, the elementary representation $\pi(l;c_1,c_2)$ contains a finite dimensional subrepresentation if and only if c_1 and c_2 are integers such that $c_1>l_1$ and $c_2<l_{n-1}$. This finite dimensional representation has as its highest weight $m_{1,n+1},m_{2,n+1},...,m_{n+1,n+1}$, where $m_{1,n+1}=c_1+1;m_{i,n+1}=l_{i-1}-i$, $i=2,3,...,n;m_{n+1,n+1}=c_2+n+1$. We denote this representation by $\Omega(c_1,l,c_2)$. Completely (or infinitesimally) irreducible representation of $U(n,1)$ is uniquely defined (up to infinitesimal equivalence) by its infinitesimal character and by the set of its irreducible representations of $U(n)$ (cf. Theorem 9.2 in Ref. 22). Infinitesimal character is given by the numbers $(l;c_1,c_2)$, which define some elementary representation of $U(n,1)$ with the same infinitesimal character. Besides the elementary irreducible and finite dimensional representations the group $U(n,1)$ has the following classes of irreducible representations[20]:

(a) The representations $D_-^{pq}(l;c_1,c_2)$ and $D_+^{pq}(l;c_1,c_2)$, $1\leqslant p\leqslant q\leqslant n, c_1>c_2$, where c_1 and c_2 are integers, such that $l_{p-1}>c_1>l_p, l_{q-1}>c_2>l_q$ (we put $l_0=\infty, l_n=-\infty$). $D_-^{pq}(l;c_1,c_2)$ and $D_+^{pq}(l;c_1,c_2)$ contain with unit multiplicity those and only those irreducible represenations of $U(n)$ for which the condition (7) and the conditions $l_{pn}\leqslant c_1, l_{qn}\leqslant c_2$ or $l_{pn}>c_1, l_{qn}>c_2$, respectively, take place.

(b) The representations $\tilde{D}_-^{pq}(l;c_1,c_2)$ and $\tilde{D}_+^{pq}(l;c_1,c_2)$, $1\leqslant p\leqslant n-1, 1\leqslant q\leqslant n$, where $c_1=l_p$ and c_2 is an integer, such that $l_{q-1}>c_2>l_q$. $\tilde{D}_-^{pq}(l;c_1,c_2)$ and $\tilde{D}_+^{pq}(l;c_1,c_2)$ contain with unit multiplicity those and only those irreducible representations of $U(n)$ for which the condition (7) and the condition $l_{qn}\leqslant c_2$ or $l_{qn}>c_2$, respectively, take place,

(c) The representations $D^i{}_-(l;c_1,c_2)$ and $D^i{}_+(l;c_1,c_2)$, $1\leqslant i\leqslant n$, where $c_1=c_2=c$ is an integer such that $l_{i-1}>c>l_i$. $D^i{}_+(l;c_1,c_2)$ and $D^i{}_+(l;c_1,c_2)$ contain with unit multiplicity those and only those irreducible representations of $U(n)$ for which the condition (7) and the condition $l_{in}\leqslant c$ or $l_{in}>c$, respectively, take place.

Among these representations there are equivalent ones.

We shall say that the sequene of integers $a_1,a_2,...,a_k$ is contracted if $a_i=a_{i-1}-1$, $i=2,3,...,k$.

The following irreducible representations are infinitesimally equivalent to the unitary representations:

(a) the representations $\pi(l;c_1,c_2)$, if

(1) c_1 and c_2 are complex numbers such that $c_1=\bar{c}_2, c_1\neq 0$,

(2) c_1 and c_2 are real numbers for which there exist such l_k and l_s $(k,s=1,2,...,n-1)$ that $|l_k-c_1|<1, |l_s-c_2|<1$, and the sequence $l_k,l_{k+1},...,l_s$, if $c_1>c_2$, or the sequence $l_s,l_{s+1},...,l_k$, if $c_1<c_2$, is contracted,

(3) c_1 and c_2 are real numbers for which there exists an integer m such that $m>c_1>m-1$, $m>c_2>m-1$;

(b) the representations $D^{ij}{}_+(l;c_1,c_2)$, if

(1) $i=j$,

(2) the sequence $c_1,l_i,l_{i+1},...,l_{j-1}$ is contracted;

(c) the representations $D^{ij}{}_-(l;c_1,c_2)$, if

(1) $i=j$,

(2) the sequence $l_i,l_{i+1},...,l_{j-1},c_2$ is contracted;

(d) the representations $\tilde{D}^{ij}{}_+(l;c_1,c_2)$, if

(1) $i<j$ and the sequence $l_i,l_{i+1},...,l_{j-1},c_2$ is contracted,

(2) $i\geqslant j$ and the sequence $l_j,l_{j+1},...,l_i$ is contracted;

(e) the representations $\tilde{D}^{ij}{}_-(l;c_1,c_2)$, if

(1) $i<j$ and the sequence $l_i,l_{i+1},...,l_{j-1}$ is contracted,

(2) $i\geqslant j$ and the sequence $c_2,l_j,l_{j+1},...,l_i$ is contracted;

(f) all the representations $D^i{}_+(l;c_1,c_2)$ and $D^i{}_-(l;c_1,c_2)$.

All these representations are unitary if the orthonormal bases in their spaces are chosen in such a way that MEs of infinitesimal operators are given by analytic continuation of the Gel'fand–Zetlin MEs of the infinitesimal operators of finite dimensional representations of the group $U(n,1)$ [or of the group $U(n+1)$], i.e., if the infinitesimal operator

$E_{n+1,n+1}$ is given by formula (10) and the infinitesimal operators $E_{n,n+1}, E_{n+1,n}$ are given by

$$E_{n,n+1}|m_n,\alpha\rangle = \sum_{i=1}^{n} [-(l_{sn}-c_1)(l_{sn}-c_2)]^{1/2} \times \omega_s(l,m_n,\alpha)|m_n^{+s},\alpha\rangle, \quad (12)$$

$$E_{n+1,n}|m_n,\alpha\rangle = \sum_{i=1}^{n} [-(l_{sn}-c_1-1)(l_{sn}-c_2-1)]^{1/2} \times \omega_s(l,m_n^{-s},\alpha)|m_n^{-s},\alpha\rangle, \quad (13)$$

where ω_s is defined by (11). The representations of (a1) belong to the principal unitary series. The represenations of (a2) and of (a3) form the complementary series. The representations of (b1) and of (c1) are the discrete series of square integrable representations.

VI. OPERATORS OF THE TRANSITION TO UNITARY REPRESENTATIONS

Since irreducible representations of U(n) are contained in elementary representations of U(n,1) with zero or unit multiplicity, it is seen from Shur's lemma and definition (6) of an intertwining operator $\Pi^{l,c_1,c_2}_{l',c_1',c_2'}$ that

$$\langle m_n,\alpha|\Pi^{l,c_1,c_2}_{l',c_1',c_2'}|m_n',\alpha'\rangle = \delta_{m_n,m_n'}\delta_{\alpha\alpha'}\lambda_{m_n}. \quad (14)$$

Thus, an intertwining operator of U(n,1) is defined by the numbers λ_{m_n}. All intertwining operators of the group U(n,1) are found in Ref. 18. In this paper we consider those which correspond to the nonidentical element w of the Weyl group W. If the elementary representation $\pi(l;c_1,c_2)$ is irreducible then for $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$

$$\lambda_{m_n} = za_{m_n}(l;c_1,c_2)/a_{m_n}(l;c_2,c_1), \quad (15)$$

where $a_{m_n}(l;c_1,c_2)$ is given as

$$a_{m_n}(l;c_1,c_2) = \prod_{r=1}^{n-1}\prod_{\sigma=l_r+1}^{l_{rn}-1}(\sigma-c_2)\prod_{\alpha=l_{sn}}^{l_{n-1,n}-1}(\alpha-c_1). \quad (16)$$

Let us define the numbers λ_{m_n} for $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$ when the representation $\pi(l;c_1,c_2)$ is reducible.

Fix l and c_1+c_2. Put $q=c_1+c_2$. Then λ_{m_n} and $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$ are functions of c_1. Moreover, they are analytic in the domain which consists of points of irreducibility of the representation $\pi(l;c_1,c_2)$ and can be meromorphically continued to the whole of the complex plane. Let $\overline{\Pi}^{l,c_1,c_2}_{l,c_2,c_1}$ be the continuation of $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$. For fixed l and c_1+c_2 the function $\overline{\Pi}^{l,c_1,c_2}_{l,c_2,c_1}$ of variable c_1 is regular [and is an intertwining operator for $\pi(l;c_1,c_2)$ and $\pi(l;c_2,c_1)$] at all points of the complex plane except integral points for which one of the following conditions is fulfilled:

(a) $l_{i-1}>c_1>l_i; l_{j-1}>c_2>l_j; i=1,2,...,n-1$; $j=1,2,...,n$; $i\neq j$ (we put $l_0=\infty, l_n=-\infty$);

(b) $l_{i-1}>c_2>c_1>l_i, i=1,2,...,n$;

(c) $c_1=l_i,\ i=1,2,...,n-1; l_{n-1}>c_2$;

(d) $c_2=l_i,\ i=1,2,...,n-1;\ l_{j-1}>c_1>l_j$, $j=1,2,...,n-1$.

At all these points except for the points c_1, for which $l_{j-1}>c_1>l_j, j=1,2,...,n-1$, the function $\overline{\Pi}^{l,c_1,c_2}_{l,c_2,c_1}$ has simple poles. At integral points c_1 for which $l_{j-1}>c_1>l_j$, $j=1,2,...,n-1$, it has poles of the second order. At every point $c_1=c_1^0$, in which $\overline{\Pi}^{l,c_1,c_2}_{l,c_2,c_1}$ has a simple pole (a pole of second order, respectively), the residue of $\overline{\Pi}^{l,c_1,c_2}_{l,c_2,c_1}$ of the first (second, respectively) order at $c_1=c_1^0$ is an intertwining operator of $\pi(l;c_1^0,c_2^0)$ and $\pi(l;c_2^0,c_1^0)$, $c_2^0=q-c_1^0$. Denote this residue by $\Pi^{l,c_1^0,c_2^0}_{l,c_2^0,c_1^0}$. It is clear that the numbers λ_{m_n} for $\Pi^{l,c_1^0,c_2^0}_{l,c_2^0,c_1^0}$ are defined to be residues at $c_1=c_1^0$ of the numbers λ_{m_n} for $\overline{\Pi}^{l,c_1,c_2}_{l,c_2,c_1}$. Some numbers λ_{m_n} for $\Pi^{l,c_1^0,c_2^0}_{l,c_2^0,c_1^0}$ are equal to 0. This means that the operator $\Pi^{l,c_1^0,c_2^0}_{l,c_2^0,c_1^0}$ vanishes on some subspace. Thus, we have defined intertwining operators for all pairs $\pi(l;c_1,c_2)$ and $\pi(l;c_2,c_1)$. We denote any of them by $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$. According to Proposition 3 in Ref. 18 the closure of the range of the noninvertible operator $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$ is invariant under the representation $\pi(l;c_2,c_1)$.

Above it was pointed out that unitarizable irreducible representation is unitary if an orthonormal basis is chosen to be such that the infinitesimal operators $E_{n,n+1}$ and $E_{n+1,n}$ are given in it by the relations (12) and (13). This orthonormal basis will be called the standard basis. Its elements will be denoted as $|m_n,\alpha\rangle_s$.

Theorem 3: Let $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$ be an intertwining operator for $\pi(l;c_1,c_2)$ and $\pi(l;c_2,c_1)$, and let H be the closure of the range of $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$ [H coincides with the space of the representation $\pi(l;c_2,c_1)$ if $\Pi^{l,c_1,c_2}_{l,c_2,c_1}$ is invertible]. If the restriction of $\pi(l;c_2,c_1)$ onto H is unitarizable or finite dimensional representation, then the transition operator A which transforms the canonical basis $|m_n,\alpha\rangle$ to the standard basis $|m_n,\alpha\rangle_s$ of H is given by

$$|m_n,\alpha\rangle_s = A\,|m_n,\alpha\rangle = (\sqrt{-1})^a\langle m_n,\alpha| \times \Pi^{l,c_1,c_2}_{l,c_2,c_1}|m_n,\alpha\rangle^{1/2}|m_n,\alpha\rangle,$$

where $a=-\Sigma_{r=1}^{n}l_{rn}$.

The proof of theorem consists in direct evaluation of MEs of the operator A using the relations (8), (9), (12), and (13). We introduce the notation

$$\mu^{l,c_1,c_2}_{m_n} = (\sqrt{-1})^a\langle m_n,\alpha|\Pi^{l,c_1,c_2}_{l,c_2,c_1}|m_n,\alpha\rangle^{1/2}. \quad (17)$$

To obtain MEs of unitary representations from ME of the elementary representations we need only the ratios $\mu_{m_n'}/\mu_{m_n}$. Really, if $\langle m_n,\alpha|T|m_n',\alpha'\rangle$ is a ME of some operator T of the representation $\pi(l;c_2,c_1)$ and ${}_s\langle m_n,\alpha|T|m_n',\alpha'\rangle_s$ is the same ME in the standard basis, then

$${}_s\langle m_n,\alpha|T|m_n',\alpha'\rangle_s = \frac{\mu^{l,c_1,c_2}_{m_n'}}{\mu^{l,c_1,c_2}_{m_n}}\langle m_n,\alpha|T|m_n',\alpha'\rangle. \quad (18)$$

Theorem 3 implies that

$$\frac{\mu^{l,c_1,c_2}_{m_n'}}{\mu^{l,c_1,c_2}_{m_n}} = \prod_{r=1}^{n}\left(\frac{\Gamma(c_1+1-l_{rn}')\Gamma(l_{rn}'-c_2)}{\Gamma(c_1+1-l_{rn})\Gamma(l_{rn}-c_2)}\right)^{1/2}. \quad (19)$$

The ratios $\Gamma(-z_1)/\Gamma(-z_2)$ with nonnegative integers z_i

and z_2 can appear in (19). In this case $\Gamma(-z_1)/\Gamma(-z_2)$ must be replaced by $(-1)^{z_1+z_2}\Gamma(1+z_2)/\Gamma(1+z_1)$.

The relations (18) and (19) will be used to obtain MEs of unitary representations of U(n,1).

VII. MATRIX ELEMENTS OF THE ELEMENTARY REPRESENTATIONS OF U(n, 1)

Any element g of U(n,1) can be decomposed uniquely as

$$g = h a_{n+1}(\varphi_{n+1}) b_{n+1}(\eta) \tilde{h}, \tag{20}$$

where $a_{n+1}(\varphi_{n+1})$ and $b_{n+1}(\eta)$ are the matrices

$$a_{n+1}(\varphi_{n+1}) = \begin{pmatrix} 1 & & 0 & \\ & \ddots & & \\ 0 & & & \\ & & 1 & 0 \\ & & 0 & e^{-i\varphi_{n+1}} \end{pmatrix}; \quad b_{n+1}(\eta) = \begin{pmatrix} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ 0 & & \cosh\eta & \sinh\eta \\ & & \sinh\eta & \cosh\eta \end{pmatrix}, \tag{21}$$

h is the most general element of the subgroup U(n), and $\tilde{h}$ is a special element of U(n). Namely, $\tilde{h}$ is defined as a product

$$\tilde{h} = a_1(\varphi_1) \prod_{r=0}^{n-2} a_{n-r}(\varphi_{n-r}) c_{n-r}(\theta_{n-r}), \tag{22}$$

where $c_{n-r}(\theta_{n-r})$ is real matrix of the rotation in the plane of $(n-r-1)$th and $(n-r)$th real coordindates, θ_{n-r} be the angle of rotation. In the relations (20)–(22)

$$0 \leqslant \varphi_i < 2\pi, \quad 1 \leqslant i \leqslant n+1; \quad 0 \leqslant \theta_j \leqslant \frac{\pi}{2}, \quad 2 \leqslant j \leqslant n; \quad 0 \leqslant \eta < \infty. \tag{23}$$

The parametrization (20)–(23) can be obtained if one carries out reasoning analogous to that of Ref. 23.

According to the decomposition (20) we can represent U(n,1) as $U(n,1) = U(n)X_{n+1}$, where X_{n+1} is a corresponding coset space. The decomposition $U(n,1) = U(n)X_{n+1}$ shows a close relation of the parameters $\{\varphi_i, \theta_j, \eta\}$ with the coordinate system on "hypersphere" S in the complex space C^{n+1}, for which the group U(n,1) is transitive.[24] It follows from this relation that an invariant measure on U(n,1) is

$$dg = \frac{n!}{2\pi^{n+1}} dh\, d\mu(x_{n+1}) = \frac{n!}{2\pi^{n+1}} \prod_{k=2}^{n} \sin^{2k-3}\theta_k \cos\theta_k \sinh^{2n-1}\eta \cosh\eta d\eta \prod_{r=1}^{n+1} d\varphi_r\, dh, \tag{24}$$

where dh is an invariant measure on U(n) such that $\int_{U(n)} dh = 1$.

To find MEs of elementary representations of U(n,1) it is enough to find MEs of the operators $T[a_{n+1}(\varphi_{n+1})]$ and $T[b_{n+1}(\eta)]$, which correspond to one-parameter subgroups $a_{n+1}(\varphi_{n+1})$ and $b_{n+1}(\eta)$, and MEs of representations of U(n). The latter will be found in Sec. IX. Now we begin to derive MEs of $T[a_{n+1}(\varphi_{n+1})]$ and $T[b_{n+1}(\eta)]$. Firstly we shall consider the irreducible elementary representations $\pi(l;c_1,c_2)$, for which c_1 and c_2 are not integers. Sometimes we shall consider the new basis $|m_n,\alpha\rangle'$ instead of the basis $|m_n,\alpha\rangle$, namely,

$$|m_n,\alpha\rangle' = \lambda(m_n,\alpha)|m_n,\alpha\rangle, \tag{25}$$

$$\lambda(m_n,\alpha) = \left[\prod_{i>j}(l_{jn}-l_{i-1}-1)!\prod_{i\leqslant j}(l_{i-1}-l_{jn})!\prod_{j=1}^{n}\Gamma(l_{jn}-c_2)\prod_{j=1}^{n}\Gamma(c_1-l_{jn}+1)\right]^{1/2}\prod_{i<j}(l_{in}-l_{jn})^{-1/2}$$
$$\times\prod_{k=2}^{n}\left[\prod_{i>j}(l_{j,k-1}-l_{ik}-1)!\prod_{i\leqslant j}(l_{ik}-l_{j,k-1})!\right]^{1/2}\prod_{i<j}(l_{i,k-1}-l_{j,k-1})^{-1/2}\mu_{m_n}^{l,c_2,c_1}, \tag{26}$$

where $\mu_{m_n}^{l,c_2,c_1}$ is defined by (17). This transformation is analogous to the transformation (22) of Ref. 25. Instead of (8) and (9) we now have

$$E_{n,n+1}|m_n,\alpha\rangle' = \sum_{s=1}^{n} \frac{\Pi_{j\leqslant s}(l_{j-1}-l_{sn})\Pi_{j<s}(l_{j,n-1}-l_{sn}-1)(c_1-l_{sn})}{\Pi_{j<s}(l_{jn}-l_{sn})\Pi_{j>s}(l_{sn}-l_{jn})}|m_n^{+s},\alpha\rangle', \tag{27}$$

$$E_{n+1,n}|m_n,\alpha\rangle' = \sum_{s=1}^{n} \frac{\Pi_{j>s}(l_{sn}-l_{j-1}-1)\Pi_{j\geqslant s}(l_{sn}-l_{j,n-1})(l_{sn}-c_2-1)}{\Pi_{j<s}(l_{jn}-l_{sn})\Pi_{j>s}(l_{sn}-l_{jn})}|m_n^{-s},\alpha\rangle'. \tag{28}$$

Infinitesimal operators of U(n) in the basis $|m_n,\alpha\rangle'$ are given by the relations (7)–(30) of Ref. 25.

For convenience we shall denote the space $L_\delta^2(K,H_\delta)$ by H. According to Corollary 4.4.5.17 in Ref. 7 every vector of dH is an analytic vector of the representation $\pi(l;c_1,c_2)$. If H^ω is a set of analytic vectors of H, then $H^\omega = \cup_{x>0} H_x^\omega$, where H_x^ω is defined as follows.[26] Let H^∞ be a space of infinitely differentiable vectors of the representation $\pi(l;c_1,c_2)$. A countable family of the seminorms ρ_m,

$$\rho_m(f) = \sup_{1 \leqslant j_k \leqslant d} \|\pi(X_{j_1} X_{j_2} \cdots X_{j_m}) f\|,$$

where d is a dimension of U$(n,1)$ and X_{j_k} are basis elements of u$(n,1)$, is constructed in H^{∞}. H_x^{ω} is a set of functions from H^{∞}, for which

$$\sum_{m=0}^{\infty} \frac{s^m}{m!} \rho_m(f) < \infty$$

for any s, $0 < s < x$.

If $|m_n,\alpha\rangle$ is a fixed element of the canonical basis in H, then we have $|m_n,\alpha\rangle \in H_x^{\omega}$ for some x, $x > 0$.

Lemma: If some element $|m_n,\alpha\rangle$ of the canonical basis of the space H of irreducible elementary representation $\pi(l;c_1,c_2)$ belongs to $H_{x_0}^{\omega}$, then all elements of this basis belong to $H_{x_0}^{\omega}$.

Proof: The representation $d\pi(l;c_1,c_2)$ of the algebra u$(n,1)$ define in an obvious way the representation of the Lie algebra $[\mathrm{u}(n,1)]^c$, the complexification of u$(n,1)$. Let $|m_n^0,\alpha^0\rangle \in H_{x_0}^{\omega}$. If V is a subspace of H, in which UIR of U(n) is realized, then according to results of Ref. 27 the universal enveloping algebra Ω of $[\mathrm{u}(n,1)]^c$ contains an element which transforms $|m_n^0,\alpha^0\rangle$ into V. Let a and b be two elements of V. According to the Bernside theorem on irreducible matrix algebras,[28] there exists in Ω an element which transforms a into b. Therefore, for any element $|m_n',\alpha'\rangle$ of the canonical basis of H there exists in Ω an element which transforms $|m_n^0,\alpha^0\rangle$ into $|m_n',\alpha'\rangle$. According to Proposition 2.1 in Ref. 26 every operator X of the representation $\pi(l;c_1,c_2)$ of the algebra u$(n,1)$ transforms H_x^{ω} into H_x^{ω}. It is obvious that this fact takes place for any element from Ω. But Ω contains elements which transfer $|m_n^0,\alpha^0\rangle$ into any element of the canonical basis, so Lemma is proved.

Let us consider the complexification $[\mathrm{U}(n,1)]^c$ of the group U$(n,1)$. According to Propositions 2.2 and 2.3 in Ref. 26 every representation $\pi(l;c_1,c_2)$ of U$(n,1)$ can be continued to the local representation of the group $[\mathrm{U}(n,1)]^c$. Notice that statements of Ref. 26 are formulated for unitary representations. However, they can be easily extended to a set of nonunitary representations which contains all elementary representations.[7]

Let $\pi(l;c_1,c_2)$ be an elementary representation of U$(n,1)$. UIR of U(n) will be called the admissible representation with respect to $\pi(l;c_1,c_2)$, if it is contained in $\pi(l;c_1,c_2)$. Highest weights of admissible representations will be called the admissible highest weights.

Proposition: Let $\pi(l;c_1,c_2)$ be an elementary representation of U$(n,1)$ for which none of the numbers c_1 and c_2 is integral. Continue the representation $\pi(l;c_1,c_2)$ to a local representation of the group $[\mathrm{U}(n,1)]^c$. There exists $t_0, t_0 > 0$, such that for $|t| < t_o, t \in \mathbb{C}$, the operator $\exp(tE_{n,n+1})$, which corresponds to the one parameter subgroup of $[\mathrm{U}(n,1)]^c$, acts upon the elements $|m_n,\alpha\rangle'$ of the space of the representations $\pi(l;c_1,c_2)$ as

$$\exp(tE_{n,n+1})|m_n,\alpha\rangle' = \sum_{m_n'} A(m_n',m_n) t^{k_n' - k_n} |m_n',\alpha\rangle'. \tag{29}$$

Here the summation is over all admissible highest weights $m_n' = (m_{1n}', m_{2n}', \ldots, m_{nn}')$ of U(n), for which $m_{in}' \geqslant m_{in}$, $i = 1,2,\ldots,n$; $k_n = \Sigma_{i=1}^{n} m_{in}$, $k_n' = \Sigma_{i=1}^{n} m_{in}'$; the multiplier $A(m_n',m_n)$ is defined by

$$A(m_n',m_n) = \prod_{i \leqslant j} \frac{(l_{i-1} - l_{jn})!}{(l_{i-1} - l_{jn}')!} \prod_{r=1}^{n} \frac{\Gamma(c_1 - l_{rn} + 1)}{\Gamma(c_1 - l_{rn}' + 1)} \prod_{i<j} \frac{(l_{i,n-1} - l_{jn} - 1)!}{(l_{i,n-1} - l_{jn}' - 1)!} \frac{\Pi_{i<j}(l_{in} - l_{jn}' - 1)!}{\Pi_{i \leqslant j}(l_{in}' - l_{jn})!} \prod_{i<j} (l_{in}' - l_{jn}'). \tag{30}$$

If $|t| < t_0, t \in \mathbb{C}$, the operator $\exp(tE_{n+1,n})$ acts on the elements $|m_n,\alpha\rangle'$ as

$$\exp(tE_{n+1,n})|m_n,\alpha\rangle' = \sum_{m_n''} B(m_n'',m_n) t^{k_n - k_n''} |m_n'',\alpha\rangle'. \tag{31}$$

Here the summation is over all the admissible highest weights $m_n'' = (m_{1n}'', m_{2n}'', \ldots, m_{nn}'')$ of U(n) for which $m_{in}'' \leqslant m_{in}$, $i = 1,2,\ldots,n$, and the multiplier $B(m_n'',m_n)$ is defined as

$$B(m_n'',m_n) = \prod_{i<j} \frac{(l_{in} - l_{j-1} - 1)!}{(l_{in}'' - l_{j-1} - 1)!} \prod_{i=1}^{n} \frac{\Gamma(l_{in} - c_2)}{\Gamma(l_{in}'' - c_2)} \prod_{i \leqslant j} \frac{(l_{in} - l_{j,n-1})!}{(l_{in}'' - l_{j,n-1})!} \frac{\Pi_{i<j}(l_{in}'' - l_{jn} - 1)!}{\Pi_{i \leqslant j}(l_{in} - l_{jn}'')!} \prod_{i<j} (l_{in}'' - l_{jn}''). \tag{32}$$

The proof is carried out in the same way as that of Theorem 1 in Ref. 25 and we omit it.

Let us remark that the operators $\exp(tE_{n,n+1})$ and $\exp(tE_{n+1,n})$ do not change α.

Theorem 4: The operators $T^{l,c_1,c_2}[a_{n+1}(\varphi)]$ and $T^{l,c_1,c_2}[b_{n+1}(\eta)]$ of the elementary representation $\pi(l;c_1,c_2)$ of U$(n,1)$, which correspond to the one-parameter subgroups $a_{n+1}(\varphi)$ and $b_{n+1}(\eta)$, in the canonical basis $|m_n,\alpha\rangle$ are given by

$$T^{l,c_1,c_2}[a_{n+1}(\varphi)]|m_n,\alpha\rangle = e^{i\varphi(k_n - k_{n+1})}|m_n,\alpha\rangle, \tag{33}$$

where

$$k_{n+1}=\sum_{i=1}^{n-1} l_i + c_1 + c_2 + \frac{(n+1)(n+2)}{2};$$

$$T^{1,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle$$
$$=\sum_{m'_n}\left[\left(\prod_{i<j}(l_{in}-l_{jn})(l'_{in}-l'_{jn})\frac{\Pi_{i<j}(l_{i,n-1}-l'_{jn}-1)!(l_{i,n-1}-l_{jn}-1)!(l_i-l'_{jn})!(l_i-l_{jn})!}{\Pi_{j\leqslant i}(l'_{jn}-l_{i,n-1})!(l_{jn}-l_{i,n-1})!(l'_{jn}-l_i-1)!(l_{jn}-l_i-1)!}\right)^{1/2}\right.$$
$$\times(\cosh\eta)^{k_n+k'_n-k_{n-1}-k_n}\sum_{m''_n}\prod_{i=1}^{n}\frac{\Gamma(c_1-l_{in}+1)\Gamma(l''_{in}-c_2)}{\Gamma(c_1-l''_{in}+1)\Gamma(l'_{in}-c_2)}\frac{\Pi_{i\leqslant j}(l''_{in}-l_j-1)!(l''_{in}-l_{j,n-1})!}{\Pi_{i\leqslant j}(l''_{in}-l'_{jn})!}$$
$$\left.\times\prod_{i<j}\frac{(l'_{in}-l''_{jn}-1)!}{(l_i-l''_{jn})!(l_{i,n-1}-l''_{jn}-1)!}\frac{\Pi_{i<j}(l_{in}-l''_{jn}-1)!(l''_{in}-l''_{jn})}{\Pi_{i\leqslant j}(l''_{in}-l_{jn})!}(\sinh\eta)^{2k''_n-k_n-k'_n}\right]|m'_n,\alpha\rangle. \tag{34}$$

Here the first sum is over all admissible highest weights m'_n of U(n) and the second sum is over the admissible highest weights m''_n of U(n), for which $m''_{in}\geqslant\max\{m_{in},m'_{in}\}$, $i=1,2,\ldots,n$. The operator $T^{1,c_1,c_2}[b_{n+1}(\eta)]$ can also be given by the formula

$$T^{1,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle$$
$$=\sum_{m'_n}\left[\left(\prod_{i<j}(l_{in}-l_{jn})(l'_{in}-l'_{jn})\frac{\Pi_{i\leqslant j}(l'_{in}-l_{j,n-1})!(l_{in}-l_{j,n-1})!(l'_{in}-l_j-1)!(l_{in}-l_j-1)!}{\Pi_{i<j}(l_{i,n-1}-l'_{jn}-1)!(l_{i,n-1}-l_{jn}-1)!(l_i-l'_{jn})!(l_i-l_{jn})!}\right)^{1/2}\right.$$
$$\times(\cosh\eta)^{k_{n-1}+k_{n-1}-k'_n-k_n}\sum_{m''_n}\prod_{j=1}^{n}\frac{\Gamma(c_1-l''_{jn}+1)\Gamma(l_{jn}-c_2)}{\Gamma(c_1-l'_{jn}+1)\Gamma(l''_{jn}-c_2)}\frac{\Pi_{i<j}(l_i-l''_{jn})!(l''_{in}-l_{jn}-1)!(l_{i,n-1}-l''_{jn}-1)!}{\Pi_{i\leqslant j}(l''_{in}-l_j-1)!(l''_{in}-l_{j,n-1})!\Pi_{i\leqslant j}(l_{in}-l''_{jn})!}$$
$$\left.\times\frac{\Pi_{i<j}(l''_{in}-l'_{jn}-1)!(l''_{in}-l''_{jn})}{\Pi_{i\leqslant j}(l'_{in}-l''_{jn})!}(\sinh\eta)^{k_n+k'_n-2k''_n}\right]|m'_n,\alpha\rangle. \tag{35}$$

Here the first sum is the same as in (34) and the second sum is over the admissible highest weights m''_n of U(n) for which $m''_{in}\leqslant\min(m_{in},m'_{in})$, $i=1,2,\ldots,n$. The ratios $\Gamma(-z_1)/\Gamma(-z_2)$ with nonnegative integers z_1 and z_2 can appear in (34) and (35). In this case $\Gamma(-z_1)/\Gamma(-z_2)$ must be replaced by $(-1)^{z_1+z_2}\Gamma(1+z_2)/\Gamma(1+z_1)$.

Proof: Let $\pi(l;c_1,c_2)$ be an elementary representation, for which none of the numbers c_1 and c_2 is integral. Continue the representation $\pi(l;c_1,c_2)$ of U(n,1) to the local representation of $[\mathrm{U}(n,1)]^c$. The element $b_{n+1}(\eta)$ of U(n,1) can be represented as

$$b_{n+1}(\eta)=\begin{pmatrix}1 & & 0 & & \\ & \ddots & & & \\ & & 1 & & \\ 0 & & & 1 & 0\\ & & & \tanh\eta & 1\end{pmatrix}\begin{pmatrix}1 & & & 0 & \\ & \ddots & & & \\ & & 1 & & \\ 0 & & & \cosh\eta & 0\\ & & & 0 & \cosh^{-1}\eta\end{pmatrix}\begin{pmatrix}1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ 0 & & & 1 & \tanh\eta\\ & & & 0 & 1\end{pmatrix}$$
$$\equiv b_+(\eta)b_0(\eta)b_-(\eta), \tag{36}$$

or as

$$b_{n+1}(\eta)=b_-(\eta)b_0^{-1}(\eta)b_+(\eta). \tag{37}$$

The elements $b_-(\eta)$, $b_0(\eta)$ and $b_+(\eta)$ belong to $[\mathrm{U}(n,1)]^c$ and do not belong to U(n,1). According to Proposition 2.2 and Proposition 2.3 in Ref. 26, and Lemma proved above, there exists η'_0, $\eta'_0>0$, such that

$$T^{1,c_1,c_2}[b_+(\eta)]T^{1,c_1,c_2}[b_0(\eta)]T^{1,c_1,c_2}[b_-(\eta)]|m_n,\alpha\rangle'=T^{1,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle', \tag{38}$$
$$T^{1,c_1,c_2}[b_-(\eta)]T^{1,c_1,c_2}[b_0^{-1}(\eta)]T^{1,c_1,c_2}[b_+(\eta)]|m_n,\alpha\rangle'=T^{1,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle' \tag{39}$$

for $\eta<\eta'_0$. Similarly, there exists η''_0, $\eta''_0>0$, such that relations (29)–(32) can be applied to the operators $T^{1,c_1,c_2}[b_-(\eta)]$ and $T^{1,c_1,c_2}[b_+(\eta)]$ if $\eta<\eta''_0$. Hence, the relations (29) and (31) can be applied to the decompositions (38) and (39) if $\eta<\eta_0\equiv\min\{\eta'_0,\eta''_0\}$. If we take into account the relations (25) and (26) this gives us (34) and (35) for $\eta<\eta_0$. Let us prove (34) and (35) for $\eta\geqslant\eta_0$. The relation (34) gives us MEs of the operator $T^{1,c_1,c_2}[b_{n+1}(\eta)]$ at any η iff at $\eta=0$ (34) gives the unit operator and the relations

$$\frac{d}{d\eta}\langle m'_n,\alpha'|T^{1,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle=\sum_{m''_n,\alpha''}\langle m'_n,\alpha'|E_{n,n+1}+E_{n+1,n}|m''_n,\alpha''\rangle\langle m''_n,\alpha''|T^{1,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle \tag{40}$$

take place[25] if MEs of $T^{1,c_1,c_2}[b_{n+1}(\eta)]$ are taken from (34). According to (9) and (10) the sum in (40) is finite. Due to the fact that this sum is finite and to properties of the infinite sums over m''_n in (34), the relation (40) is valid for any η if it does for some interval of η. But for $\eta<\eta_0$ the relation (40) is valid since for these η the relation (34) gives us MEs of $T^{1,c_1,c_2}(b_{n+1}(\eta))$. Hence, (40) takes place for all η and the relation (34) is proved for $\pi(l;c_1,c_2)$. The relation (35) is proved in the same way if we take

infinite sum over m''_{nn} instead of m''_{1n}. Thus, Theorem is proved for all the representations $\pi(l;c_1,c_2)$ for which none of the numbers c_1 and c_2 is an integer. According to Theorem 1 MEs of the elementary representation $\pi(l;c_1,c_2)$ are entire analytic functions of c_1 at fixed l and c_1+c_2. Every infinite sum over m''_{1n} in (34) and over m''_{nn} in (35) permits an analytic continuation in c_1. An analytic continuation of (34) and (35) gives us MEs of the operator $T^{l,c_1,c_2}[b_{n+1}(\eta)]$ for any elementary representation. The theorem is proved.

Remark: An explicit form of ME (34) and the range of summing over m''_n shows that ME (34) is an infinite sum over m''_{1n} for every case when c_1 is not an integer such that $c_1 > l_1$. If c_1 is an integer such that $c_1 > l_1$ then ME (34) is an infinite sum if and only if $l_{1n} > c_1$, $l_{1n} = m_{1n} - 1$. ME (35) is an infinite sum over m''_{nn} for every case when c_2 is not an integer such that $c_2 < l_{n-1}$. If c_2 is an integer such that $c_2 < l_{n-1}$ then ME (35) is an infinite sum if and only if $l_{nn} \leqslant c_2$, $l_{nn} = m_{nn} - n$.

MEs (34) and (35) can be represented as sums of generalized hypergeometric functions ${}_{q+1}F_q$. Indeed, if we choose in (34) m''_{1n} as the summation parameter of a generalized hypergeometric series then, after simple transformations, ME (34) for $m'_{1n} \geqslant m_{1n}$ takes the form

$$\begin{aligned}
&\langle m'_n,\alpha|T^{l,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle \\
&= \left[\prod_{i<j}(l_{in}-l_{jn})(l'_{in}-l'_{jn})\frac{\Pi_{i<j}(l_{i,n-1}-l_{jn}-1)!}{\Pi_{i\leqslant j}(l_{in}-l_{j,n-1})!}\,\frac{\Pi_{i<j}(l'_{i,n-1}-l_{jn}-1)!(l_i-l'_{jn})!(l_i-l_{jn})!}{\Pi_{i\leqslant j}(l'_{in}-l_{j,n-1})!(l_{in}-l_j-1)!(l'_{in}-l_j-1)!}\right]^{1/2} \\
&\quad\times\prod_{j=1}^{n}\frac{\Gamma(c_1-l_{jn}+1)}{\Gamma(l'_{jn}-c_2)}(\sinh\eta)^{-k_n-k'_n}(\cosh\eta)^{k_n+k'_n-k_{n-1}-k_{n+1}}\sum_{m''_{2n}}\cdots\sum_{m''_{nn}}(\sinh\eta)^{2(k''_n-m''_{1n})}\prod_{1<i\leqslant j}\frac{(l''_{in}-l_{j,n-1})!}{(l''_{in}-l'_{jn})!} \\
&\quad\times\prod_{i<j}\frac{(l'_{in}-l''_{jn}-1)!(l_{in}-l''_{jn}-1)!}{(l_{i,n-1}-l''_{jn}-1)!}\,\frac{\Pi_{1<i<j}(l''_{in}-l''_{jn})\,\Pi_{1<i\leqslant j}(l''_{in}-l_j-1)!}{\Pi_{1<i\leqslant j}(l''_{in}-l_{jn})!\quad\Pi_{i<j}(l_i-l''_{jn})!}\prod_{i=2}^{n}\frac{\Gamma(l''_{in}-c_2)}{\Gamma(c_1-l''_{jn}+1)} \\
&\quad\times W^{l,c_1,c_2}_{m''_n,m'_n}(\eta)\,{}_{3n-1}F_{3n-2}\left[\begin{matrix}l'_{1n}-c_1,l'_{1n}-c_2,\{l'_{1n}-l_i\}_1^{n-1},\{l'_{1n}-l_{j,n-1}+1\}_1^{n-1},\{l'_{1n}-l''_{jn}+1\}_2^{n}\\ \{l'_{1n}-l''_{jn}\}_2^{n},\{l'_{1n}-l'_{jn}+1\}_2^{n},\{l'_{1n}-l_{jn}+1\}_1^{n}\end{matrix}\,\middle|\,-\sinh^2\eta\right],
\end{aligned} \tag{41}$$

where

$$W^{l,c_1,c_2}_{m''_n,m'_n}(\eta)=(\sinh\eta)^{2m'_{1n}}\frac{\Gamma(l'_{1n}-c_2)}{\Gamma(c_1-l'_{1n}+1)}\prod_{j=2}^{n}\frac{(l'_{1n}-l_j-1)!(l'_{1n}-l''_{jn})}{(l'_{1n}-l'_{jn})!}\,\frac{\Pi_{j=1}^{n-1}(l'_{1n}-l_{j,n-1})!}{\Pi_{j=1}^{n}(l'_{1n}-l_{jn})!}.$$

The relation (41) gives ME (34) for $m'_{1n} < m_{1n}$ if the permutation $m'_{1n} \leftrightarrow m_{1n}$ is made in the product $W^{l,c_1,c_2}_{m''_n,m'_n}(\eta)\,{}_{3n-1}F_{3n-2}$. In (41) the sums are over all the integers $m''_{2n},\ldots,m''_{nn}$, for which $m_{i-1} \geqslant m''_{in} \geqslant \max\{m_{in},m'_{in}\}$ where $m_{i-1} = l_{i-1} + i$. The notation $\{a_i\}_p^q$, $p \leqslant q$, means a set of the numbers a_p, $a_{p+1},\ldots,a_q$.

For $m'_{nn} \geqslant m_{nn}$ ME (35) can be represented as

$$\begin{aligned}
&\langle m'_n,\alpha|T^{l,c_1,c_2}[b_{n+1}(\eta)]|m_n,\alpha\rangle \\
&= \left[\prod_{i<j}(l_{in}-l_{jn})(l'_{in}-l'_{jn})\frac{\Pi_{i\leqslant j}(l'_{in}-l_{j,n-1})!(l_{in}-l_{j,n-1})!(l'_{in}-l_j-1)!(l_{in}-l_j-1)!}{\Pi_{i<j}(l_{i,n-1}-l'_{jn}-1)!(l_{i,n-1}-l_{jn}-1)!(l_i-l'_{jn})!(l_i-l_{jn})!}\right]^{1/2} \\
&\quad\times\prod_{i=1}^{n}\frac{\Gamma(l_{in}-c_2)}{\Gamma(c_1-l'_{in}+1)}(\cosh\eta)^{k_{n-1}+k_{n+1}-k_n-k'_n}(\sinh\eta)^{k_n+k'_n}\sum_{m''_{1n}}\cdots\sum_{m''_{n-1,n}}(\sinh\eta)^{2(-m''_{1n}-\cdots-m''_{n-1,n})} \\
&\quad\times\prod_{i<j}(l''_{in}-l_{jn}-1)!\frac{\Pi_{i<j<n}(l''_{in}-l''_{jn})(l_{i,n-1}-l''_{jn}-1)!(l_i-l''_{jn})!}{\Pi_{i\leqslant j<n}(l_{in}-l''_{jn})!(l'_{in}-l''_{jn})!(l''_{in}-l_j-1)!}\,\frac{\Pi_{i<j}(l''_{in}-l'_{jn}-1)!}{\Pi_{i\leqslant j}(l''_{in}-l_{j,n-1})!} \\
&\times\prod_{i=1}^{n-1}\frac{\Gamma(c_1-l''_{jn}+1)}{\Gamma(l''_{in}-c_2)}V^{l,c_1,c_2}_{m''_n,m'_n}(\eta)_{3n-1}F_{3n-2} \\
&\quad\times\left[\begin{matrix}c_1-l_{nn}+1,c_2-l_{nn}+1,\{l_i-l_{nn}+1\}_1^{n-1},\{l_{i,n-1}-l_{nn}\}_1^{n-1},\{l''_{in}-l_{nn}+1\}_1^{n-1}\\ \{l_{in}-l_{nn}+1\}_1^{n-1},\{l'_{in}-l_{nn}+1\}_1^{n},\{l''_{in}-l_{nn}\}_1^{n-1}\end{matrix}\,\middle|\,-\sinh^2\eta\right],
\end{aligned} \tag{42}$$

where

$$V^{l,c_1,c_2}_{m''_n,m'_n}(\eta)=(\sinh\eta)^{-2m_{nn}}\frac{\Gamma(c_1-l_{nn}+1)}{\Gamma(l_{nn}-c_2)}\prod_{i=1}^{n-1}\frac{(l_i-l_{nn})!(l_{i,n-1}-l_{nn}-1)!}{(l_{in}-l_{nn})!}\,\frac{\Pi_{i=1}^{n-1}(l''_{in}-l_{nn})}{\Pi_{i=1}^{n}(l'_{in}-l_{nn})!}.$$

The relation (42) gives us ME (35) also for $m'_{nn} < m_{nn}$ if the permutation $m_{nn} \leftrightarrow m'_{nn}$ is made in the product $V^{l,c_1,c_2}_{m''_n,m'_n}(\eta)\,{}_{3n-1}F_{3n-2}$. In (42) the sums are over all the integers $m''_{1n},\ldots,m''_{n-1,n}$, for which $\min\{m_{in},m'_{in}\} \geqslant m''_{in} \geqslant m_i$.

VIII. MATRIX ELEMENTS OF UNITARY AND FINITE DIMENSIONAL REPRESENTATIONS OF U(n, 1)

MEs of unitary and finite dimensional representations of U(n,1) in the standard basis will be denoted by ${}_s\langle m_n,\alpha|T|m'_n,\alpha'\rangle_s$. To find them we shall use Theorem 4 and the relations (18) and (19).

Classification of UIRs of U(n,1) is given in Sec. V. Below we use the notation of classes of UIRs introduced in Sec. V. For $n+1$ numbers $(l;c_1,c_2)\equiv(l_1,l_2,\ldots,l_{n-1},c_1,c_2)$ and for the transposition s of these numbers we define $n+1$ numbers $s(l,c_1,c_2)$. The permutation of the ith and jth numbers is denoted by s_{ij}.

Theorem 5: The unitarizable representations of (a1) and (f1) of the classification of UIRs of U(n,1) are unitary in an orthonormal basis of the space $L^2_\delta(K,H_\delta)$. Their MEs in the canonical basis are given by (33)–(35). MEs of the operator $T(b_{n+1}(\eta))$ of any other unitarizable representation, characterized by the numbers l,c_1,c_2, or of the finite dimensional representation $\Omega(c_1,l,c_2)$ in the standard basis are given by

$$
{}_s\langle m_n,\alpha|T(b_{n+1}(\eta))|m'_n,\alpha\rangle_s
$$

$$
=(-1)^\beta\frac{\mu^{l,c_1,c_2}_{m'_n}}{\mu^{l,c_1,c_2}_{m_n}}\langle m_n,\alpha|T^{l,c_1,c_2}(b_{n+1}(\eta))|m'_n,\alpha\rangle, \tag{43}
$$

where the right-hand side contains ME of the corresponding elementary representation $\pi(l;c_2,c_1)$, and $\beta=0$ for the unitarizable representations of (a2), (a3), (b1), (c1), (e1), (e2), and finite dimensional representations; $\beta=\Sigma_{r=1}^{n}(l_{rn}+l'_{rn})$ for the representations of (d1), (d2), and (c2); $\beta=\Sigma_{r=i+1}^{n}(l_{rn}+l'_{rn})$ for the representations of (b2).

Proof: According to Theorem 3 and relations (18) and (19), first of all it is necessary to find the elementary representations which contain unitarizable and finite dimensional representations as subrepresentations. The composition series (i.e., the structure) of elementary representations are described in Ref. 20. According to results of Ref. 20, the elementary representation $\pi(l';c'_2,c'_1)$ contains the unitarizable representation, characterized by the numbers $(l;c_1,c_2)$ (given in the classification of unitary representations in Sec. V), if $(l',c'_1,c'_2)=(l,c_1,c_2)$ for the representations of (a2), (a3), (b1), (c1), (e1), (e2); if $(l',c'_1,c'_2)=(s_{n,n+1}s_{in}(l,c_1,c_2)$ for the representations (b2); if $(l',c'_1,c'_2)=s_{n,n+1}s_{n+1,j-1}(l,c_1,c_2)$ for the representations of (c2) for which $i<j-1$; if $(l',c'_1,c'_2)=s_{n,n+1}s_{n+1,i}(l,c_1,c_2)$ for the representations of (c2), for which $j=i+1$; if $(l',c'_1,c'_2)=s_{n,n+1}(l,c_1,c_2)$ for the representations of (f1) and (f2). According to (18) and (19) MEs of the unitarizable representation, characterized by the numbers l,c_1,c_2, in the standard basis are

$$
{}_s\langle m_n,\alpha|T[b_{n+1}(\eta)]|m'_n,\alpha\rangle_s
$$

$$
=\frac{\mu^{l,c'_1,c'_2}_{m'_n}}{\mu^{l,c'_1,c'_2}_{m_n}}\langle m_n,\alpha|T^{l,c'_1,c'_2}[b_{n+1}(\eta)]|m'_n,\alpha\rangle, \tag{44}
$$

where MEs of the right-hand side are given by Theorem 4. Applying the relation $\Gamma(z)\Gamma(1-z)=\pi/\sin(\pi z)$ to the right-hand side of (44), after simple transformations we obtain (43) for every type of unitary representations and for finite dimensional representations. The theorem is proved.

Let us note that the relation (43) with $\beta=0$ also defines unitarized MEs of the operator $T(b_{n+1}(\eta))$ of the principal unitary series representations. This means that the operator A of Theorem 3 is unitary if the representation $\pi(l;c_1,c_2)$ belongs to the principal unitary series.

It follows from Theorem 5 and Remark in Sec. VII that for some UIRs all their MEs are expressed as finite sums. This is true for the UIRs of U(n,1) which contain only those UIRs of U(n), for which $l_{1n}\leqslant m$ or $l_{nn}\geqslant m'$ for some fixed m and m'. Theorem 5 and the Remark in Sec. VII show that MEs of finite dimensional irreducible representations are finite sums.

IX. MATRIX ELEMENTS AND CLEBSCH–GORDAN COEFFICIENTS OF U(n)

It was shown in Sec. III that MEs of UIRs of U($n+1$) can be obtained from MEs of finite dimensional representations of U(n,1) by the analytic continuation of parameters of the group U(n,1) to parameters of U($n+1$). According to results of Ref. 23, any element g, $g\in$U($n+1$), can be represented as

$$
g=ha_{n+1}(\varphi_{n+1})\beta_{n+1}(\theta_{n+1})\tilde{h}, \tag{45}
$$

where $h,\tilde{h}$, and $a_{n+1}(\varphi_{n+1})$ are the same as in (20)–(22), and

$$
\beta_{n+1}(\theta_{n+1})=\begin{pmatrix}1 & & & & 0\\ & \ddots & & & \\ & & 1 & & \\ & & & \cos\theta_{n+1} & -\sin\theta_{n+1}\\ 0 & & & \sin\theta_{n+1} & \cos\theta_{n+1}\end{pmatrix},\quad 0\leqslant\theta_{n+1}\leqslant\frac{\pi}{2}. \tag{46}
$$

Therefore, we have to continue analytically the parameter η. But the continuation $\eta\to i\theta_{n+1}$ does not transfer the matrix $b_{n+1}(\eta)$ into the matrix $\beta_{n+1}(\theta_{n+1})$. We have to continue the matrix $s^{-1}b_{n+1}(\eta)s$, where $s=\left(\begin{smallmatrix}E & 0\\ 0 & i\end{smallmatrix}\right)$, E being the unit $(n\times n)$-matrix. Really,

$$s^{-1}b_{n+1}(\eta)s=\begin{pmatrix}1 & & & & 0\\ & \ddots & & & \\ & & 1 & & \\ & & & \cosh\eta & i\sinh\eta\\ 0 & & & -i\sinh\eta & \cosh\eta\end{pmatrix}\xrightarrow{\eta\to i\theta}\begin{pmatrix}1 & & & & 0\\ & \ddots & & & \\ & & 1 & & \\ & & & \cos\theta & -\sin\theta\\ 0 & & & \sin\theta & \cos\theta\end{pmatrix}. \tag{47}$$

Hence, we shall obtain unitarized MEs of the operator $T(\beta_{n+1}(\theta_{n+1}))$ of UIR of U$(n+1)$ if we do the analytic continuation $\eta\to i\theta$ in MEs of the operator $T(s^{-1}b_{n+1}(\eta)s)$ of a finite dimensional representation of U$(n,1)$, provided these MEs are written in the standard basis. As a result we have

$$\begin{aligned}&{}_s\langle\tilde m_n,\alpha|T(\beta_{n+1}(\theta))|m_n,\alpha\rangle_s\\ &=\left[\prod_{i<j}(l_{in}-l_{jn})(\tilde l_{in}-\tilde l_{jn})\frac{\Pi_{i<j}(l_{i,n-1}-\tilde l_{jn}-1)!(l_{i,n-1}-l_{jn}-1)!\Pi_{i\leqslant j}(l_{i,n+1}-l_{jn})!(l_{i,n+1}-\tilde l_{jn})!}{\Pi_{i<j}(\tilde l_{in}-l_{j,n+1}-1)!(l_{in}-l_{j,n+1}-1)!\Pi_{i\leqslant j}(l_{in}-l_{j,n-1})!(\tilde l_{in}-l_{j,n-1})!}\right]^{1/2}\\ &\times(\cos\theta)^{k_n+\tilde k_n-k_{n-1}-k_{n+1}}\sum_{m'_{1n},\dots,m'_{nn}}(-1)^{k'_n-k_n}(\sin\theta)^{2k'_n-k_n-\tilde k_n}\\ &\times\frac{\Pi_{i<j}(l'_{in}-l_{j,n+1}-1)!\Pi_{i\leqslant j}(l'_{in}-l_{j,n-1})!\Pi_{i<j}(\tilde l_{in}-l'_{jn}-1)!(l_{in}-l'_{jn}-1)!}{\Pi_{i\leqslant j}(l_{i,n+1}-l'_{jn})!\Pi_{i<j}(l_{i,n-1}-l'_{jn}-1)!\Pi_{i\leqslant j}(l'_{in}-\tilde l_{jn})!(l'_{in}-l_{jn})!}\prod_{i<j}(l'_{in}-l'_{jn}).\end{aligned} \tag{48}$$

Here $l_{1,n+1},l_{2,n+1},\dots,l_{n+1,n+1}$ are the numbers which are defined by the highest weight $(m_{1,n+1},m_{2,n+1},\dots,m_{n+1,n+1})$ of the representation by the relation $l_{i,n+1}=m_{i,n+1}-i$. The sum is over all m'_n, for which $m_{i,n+1}\geqslant m'_{i,n}\geqslant\max(m_{in},\tilde m_{in})$. The relation (48) is obtained under the condition that for the right-hand side of (43) the relation (34) was used. The relation (35) for the right-hand side of (43) leads to

$$\begin{aligned}&{}_s\langle\tilde m_n,\alpha|T[\beta_{n+1}(\theta)]|m_n,\alpha\rangle_s\\ &=\left[\prod_{i<j}(l_{in}-l_{jn})(\tilde l_{in}-\tilde l_{jn})\frac{\Pi_{i<j}(l_{in}-l_{j,n+1}-1)!(\tilde l_{in}-l_{j,n+1}-1)!\Pi_{i\leqslant j}(l_{in}-l_{j,n-1})!(\tilde l_{in}-l_{j,n-1})!}{\Pi_{i<j}(l_{i,n-1}-l_{jn}-1)!(l_{i,n-1}-\tilde l_{jn}-1)!\Pi_{i\leqslant j}(l_{i,n+1}-\tilde l_{jn})!(l_{i,n+1}-l_{jn})!}\right]^{1/2}\\ &\times(\cos\theta)^{k_{n-1}+k_{n-1}-k_n-\tilde k_n}\sum_{m'_{1n},\dots,m'_{nn}}(-1)^{\tilde k_n-k'_n}(\sin\theta)^{-2k'_n+k_n+\tilde k_n}\\ &\times\frac{\Pi_{i\leqslant j}(l_{i,n+1}-l'_{jn})!\Pi_{i<j}(l_{i,n-1}-l'_{jn}-1)!(l'_{in}-l_{jn}-1)!(l'_{in}-\tilde l_{jn}-1)!}{\Pi_{i<j}(l'_{in}-l_{j,n+1}-1)!\Pi_{i\leqslant j}(l'_{in}-l_{j,n-1})!\Pi_{i\leqslant j}(l_{in}-l'_{jn})!(\tilde l_{in}-l'_{jn})!}\prod_{i<j}(l'_{in}-l'_{jn}).\end{aligned} \tag{49}$$

The sum is over all m'_n for which $m_{i+1,n+1}\leqslant m_{in}\leqslant\min(m_{in},\tilde m_{in})$.

The MEs (48) and (49) can be written through the generalized hypergeometric function ${}_{3n-1}F_{3n-2}$.

Now find MEs of a representation operator corresponding to an arbitrary element of U(n). According to (45)

$$\langle\bar m_{n-1},\alpha|T^{m_n}(g)|m_{n-1},\alpha\rangle=D^{m_n}_{\bar\alpha\ \alpha}{}_{m_n}(ha_nb_n\tilde h)=\sum_{\dot\alpha}D^{m_n}_{\bar\alpha\ \dot\alpha}{}_{\bar m_n}(h)D^{m_n}_{\dot\alpha\ \alpha}{}_{m_n}(a_nb_n)D^{m_n}_{\bar\alpha\ \alpha}{}_{m_n}(\tilde h). \tag{50}$$

Put $h=e$ [e is the unit of U(n)] in (50). Then

$$D^{m_n}_{\bar\alpha\ \alpha}{}_{m_n}(a_nb_n\tilde h)=A^{\bar m_n}_{m_n}(\varphi_n)d^{\bar m_n}_{m_{n-1}\,\bar m_{n-1}}(\theta)D^{m_{n-1}}_{\bar\alpha'\ \alpha'}{}_{m_{n-1}}(\tilde h), \tag{51}$$

where $A^{\bar m_n}_{m_n}(\varphi_n)$ is ME of the representation operator for $a_n=a_n(\varphi_n)$ and α' is the Gel'fand–Zetlin scheme α without the first line m_{n-2}. Since $\tilde h$ decomposes according to (22), the relation (55) gives us

$$D^{m_{n-1}}_{\bar\alpha\ \alpha}{}_{m_{n-1}}(\tilde h)=A^{m_1}(\varphi_1)\prod_{i=1}^{n-2}A^{\bar m_{n-i}}_{m_{n-i}}(\varphi_{n-i})d^{m_{n-i}}_{m_{n-i-1}\,\bar m_{n-i-1}}(\theta_{n-i}). \tag{52}$$

Thus, MEs of the representation operator corresponding to an arbitrary element of U(n), are defined by (51), (33), (52), (48), and (49).

Now we use MEs of UIRs of U(n) to derive CGCs of this group. The following integral relation is valid,

$$\begin{aligned}&\sum_\gamma\left\langle\begin{matrix}\bar m_n & \tilde m_n & m_n\\ \bar m_{n-1} & \tilde m_{n-1} & m_{n-1}\\ \vdots & \vdots & \vdots\end{matrix}\middle|\gamma\right\rangle^*\left\langle\begin{matrix}\bar m_n & \tilde m_n & m_n\\ \bar m'_{n-1} & \tilde m'_{n-1} & m'_{n-1}\\ \vdots & \vdots & \vdots\end{matrix}\middle|\gamma\right\rangle\\ &=(\dim T^{m_n})\int_{U(n)}D^{\bar m_n}_{\vdots\ \vdots\ \bar m'_n}{}_{m_n}(g)D^{\bar m_n}_{\vdots\ \vdots\ \bar m'_{n-1}}{}_{m_n}\left[(g)D^{m_n}_{\vdots\ \vdots\ m'_n}{}_{m_n}(g)\right]^*dg,\end{aligned} \tag{53}$$

where γ labels the multiple representations T^{m_n} in the tensor product of the representations T^{m_n} and T^{m_n}; $*$ means the complex conjugation.

The invariant measure dg on U(n) is defined by means of the decomposition (45) in the same way as in the case of the group U(n,1). Namely, if $g = ha_n(\varphi_n)b_n(\theta_n)\tilde{h}$, $g \in U(n)$, then

$$dg = \frac{(n-1)!}{2\pi^n} \prod_{k=2}^{n} \sin^{2k-3}\theta_k \cos\theta_k d\theta_k \prod_{r=1}^{n} d\varphi_r\, dh, \tag{54}$$

where dh is invariant measure on U($n-1$), normalized by the condition $\int_{U(n-1)} dh = 1$. Taking into account (51), (52), (54), and using the relation (53) for the subgroup U($n-1$), we have

$$\begin{aligned}
&\sum_{\gamma_n} \left\langle \begin{matrix} \bar{m}_n & \tilde{m}_n & m_n \\ \bar{m}_{n-1} & \tilde{m}_{n-1} & m_{n-1} \\ \vdots & \vdots & \vdots \end{matrix} \right| \gamma_n \Bigg\rangle^* \left\langle \begin{matrix} m_n & \tilde{m}_n & m_n \\ \bar{m}'_{n-1} & \tilde{m}'_{n-1} & m'_{n-1} \\ \vdots & \vdots & \vdots \end{matrix} \right| \gamma_n \Bigg\rangle \\
&= \frac{(n-1)!}{2\pi^n} \frac{\dim T^{m_n}}{\dim T^{m_n}} \int_0^{2\pi} d\varphi_n A^{\bar{m}_n}_{\bar{m}_n\ 1}(\varphi_n) A^{\tilde{m}_n}_{\tilde{m}_n\ 1}(\varphi_n) A^{m_n}_{m_n\ 1}(\varphi_n)^* \sum_{\gamma_{n-1}} \left\langle \begin{matrix} \bar{m}_{n-1} & \tilde{m}_{n-1} & m_{n-1} \\ \bar{m}_{n-2} & \tilde{m}_{n-2} & m_{n-2} \\ \vdots & \vdots & \vdots \end{matrix} \right| \gamma_{n-1} \Bigg\rangle^* \\
&\times \sum_{\substack{\dot{m}_n \\ \vdots}} \sum_{\substack{\dot{m}_{n-1} \\ \vdots}} \sum_{\substack{\dot{m}_{n-2} \\ \vdots}} \left\langle \begin{matrix} \bar{m}_{n-1} & \tilde{m}_{n-1} & m_{n-1} \\ \dot{m}_{n-2} & \ddot{m}_{n-2} & \hat{m}_{n-2} \\ \vdots & \vdots & \vdots \end{matrix} \right| \gamma_{n-1} \Bigg\rangle I_n \begin{Bmatrix} \bar{m}_n; \bar{m}_{n-1}; \bar{m}'_{n-1}; \dot{m}_{n-2} \\ \tilde{m}_n; \tilde{m}_{n-1}; \tilde{m}'_{n-1}; \ddot{m}_{n-2} \\ m_n; m_{n-1}; m'_{n-1}; \hat{m}_{n-2} \end{Bmatrix} \\
&\times \prod_{k=2}^{n-1} I_k \begin{Bmatrix} \bar{m}'_k; \dot{m}_{k-1}; \bar{m}'_{k-1}; \dot{m}_{k-2} \\ \tilde{m}'_k; \ddot{m}_{k-1}; \tilde{m}'_{k-1}; \ddot{m}_{k-2} \\ m'_k; \hat{m}_{k-1}; m'_{k-1}; \hat{m}_{k-2} \end{Bmatrix} \int_0^{2\pi} d\varphi_1 A^{\bar{m}'_1}(\varphi_1) A^{\tilde{m}'_1}(\varphi_1) A^{m'_1}(\varphi_1)^* \prod_{i=1}^{n-1} \int_0^{2\pi} A^{\bar{m}'_i}_{\bar{m}'_i\ 1}(\varphi_i) A^{\tilde{m}'_i}_{\tilde{m}'_i\ 1}(\varphi_i) A^{m'_i}_{\hat{m}'_i\ 1}(\varphi_i)^* d\varphi_i,
\end{aligned} \tag{55}$$

where

$$\begin{aligned}
&I_k \begin{Bmatrix} \bar{m}_k; \bar{m}_{k-1}; \bar{m}'_{k-1}; \bar{m}_{k-2} \\ \tilde{m}_k; \tilde{m}_{k-1}; \tilde{m}'_{k-1}; \tilde{m}_{k-2} \\ m_k; m_{k-1}; m'_{k-1}; m_{k-2} \end{Bmatrix} \\
&= \int_0^{\pi/2} \sin^{2k-3}\theta_k \cos\theta_k d\theta_k d^{\bar{m}_k}_{\bar{m}_k\ 1, \bar{m}'_k\ 1; \bar{m}_k\ 2}(\theta_k) d^{\tilde{m}_k}_{\tilde{m}_k\ 1, \tilde{m}'_k\ 1; \tilde{m}_k\ 2}(\theta_k) [d^{m_k}_{m_k\ 1, m'_k\ 1; m_k\ 2}(\theta_k)]^*.
\end{aligned} \tag{56}$$

Now carry out the integrations. According to (33)

$$\int_0^{2\pi} d\varphi_i A^{\bar{m}_i}_{\bar{m}_i\ 1}(\varphi_i) A^{\tilde{m}_i}_{\tilde{m}_i\ 1}(\varphi_i) A^{m_i}_{m_i\ 1}(\varphi_i) = 2\pi \delta_{0, \bar{k}_i\ 1 + \tilde{k}_i\ 1 - k_i\ 1 + k_i - \bar{k}_i - \tilde{k}_i}. \tag{57}$$

To complete the integration in (56), we represent the relations (48) and (49) as

$$\begin{aligned}
d^{m_n}_{\bar{m}_n\ 1, m_n\ 1; m_n\ 2}(\theta) = &\sum_{m'_{1,n}\ 1, \ldots, m'_n\ 1, n-1} F^{(1)}_{n-1}(m_n; \tilde{m}_{n-1}, m_{n-1}; m_{n-2}) \\
&\times (\sin\theta)^{2k'_n\ 1 - k_n\ 1 - \tilde{k}_n\ 1} (\cos\theta)^{k_n\ 1 + \tilde{k}_n\ 1 - k_n\ 2 - k_n},
\end{aligned} \tag{58}$$

$$\begin{aligned}
d^{m_n}_{\tilde{m}_n\ 1, m_n\ 1; m_n\ 2}(\theta) = &\sum_{m'_{1,n}\ 1, \ldots, m'_n\ 1, n-1} F^{(2)}_{n-1}(m_n; \tilde{m}_{n-1}, m_{n-1}; m_{n-2}) \\
&\times (\sin\theta)^{-2k'_n\ 1 + k_n\ 1 + \tilde{k}_n\ 1} (\cos\theta)^{k_n + k_n\ 2 - k_n\ 1 - \tilde{k}_n\ 1}.
\end{aligned} \tag{59}$$

Now the relation (56) can be written as

$$\begin{aligned}
&I_s \begin{Bmatrix} \bar{m}_s; \bar{m}_{s-1}; \bar{m}'_{s-1}; \bar{m}_{s-2} \\ \tilde{m}_s; \tilde{m}_{s-1}; \tilde{m}'_{s-1}; \tilde{m}_{s-2} \\ m_s; m_{s-1}; m'_{s-1}; m_{s-2} \end{Bmatrix} \\
&= \sum_{\bar{\bar{m}}_{1,s}\ 1, \ldots, \bar{\bar{m}}_s\ 1, s-1} \sum_{\tilde{\tilde{m}}_{1,s}\ 1, \ldots, \tilde{\tilde{m}}_s\ 1, s-1} \sum_{m''_{1,s}\ 1, \ldots, m''_s\ 1, s-1} Q_1 \int_0^{\pi/2} d\theta_s (\sin\theta_s)^{\mu-1} (\cos\theta_s)^{\nu-1},
\end{aligned} \tag{60}$$

where

$$Q_1 = F^{(1)}_{s-1}(\bar{m}_s; \bar{m}_{s-1}, \bar{m}'_{s-1}; \bar{m}_{s-2}) F^{(1)}_{s-1}(\tilde{m}_s; \tilde{m}_{s-1}, \tilde{m}'_{s-1}; \tilde{m}_{s-2}) F^{(1)}_{s-1}(m_s; m_{s-1}; m'_{s-1}; m_{s-2}), \tag{61}$$

$$\mu = 2(s - 1 + \bar{\bar{k}}_{s-1} + \tilde{\tilde{k}}_{s-1} + k''_{s-1}) - \bar{k}_{s-1} - \bar{k}'_{s-1} - \tilde{k}_{s-1} - \tilde{k}'_{s-1} - k_{s-1} - k'_{s-1}, \tag{62}$$

$$\nu = 2 + \bar{k}_{s-1} - \bar{k}'_{s-1} + \tilde{k}_{s-1} + \tilde{k}'_{s-1} + k_{s-1} + k'_{s-1} - \bar{k}_s - \tilde{k}_s - k_s - \bar{k}_{s-2} - \tilde{k}_{s-2} - k_{s-2}, \tag{63}$$

and the sum is the same as in (58). According to the relation 3.621(5) in Ref. 29 the integral (60) exists if $\mathrm{Re}\mu > 0$, $\mathrm{Re}\nu > 0$, and under these conditions

$$\int_0^{\pi/2} (\sin\theta)^{\mu-1}(\cos\theta)^{\nu-1}d\theta = \tfrac{1}{2}B(\mu/2,\nu/2), \tag{64}$$

where $B(x,y)$ is a beta function. It is obvious that the parameter μ in (62) is always positive. The parameter ν in (63) can be both positive and negative. If $\nu > 0$, then the integral in (60) is equal to (64) with μ and ν defined by (62) and (63), respectively.

Let $\nu \leqslant 0$. Then

$$\bar{k}_s + \tilde{k}_s + k_s + \bar{k}_{s-2} + \tilde{k}_{s-2} + k_{s-2} \geqslant \bar{k}_{s-1} + \bar{k}'_{s-1} + \tilde{k}_{s-1} + \tilde{k}'_{s-1} + k_{s-1} + k'_{s-2} + 2\,. \tag{65}$$

In this case we use d-function (59) for the integrand in (56). We have

$$I_s \begin{Bmatrix} \bar{m}_s;\bar{m}_{s-1};\bar{m}'_{s-1};\bar{m}_{s-2} \\ \tilde{m}_s;\tilde{m}_{s-1};\tilde{m}'_{s-1};\tilde{m}_{s-2} \\ m_s;m_{s-1};m'_{s-1};m_{s-2} \end{Bmatrix} = \sum_{\bar{m}_{1,2}\ldots,\bar{m}_{1,s-1}} \sum_{\tilde{m}_{1,s-1}\ldots,\tilde{m}_{s-1,s-1}} \sum_{m''_{1,s-1}\ldots,m''_{s-1,s-1}} Q_2 \int_0^{\pi/2} d\theta_s (\sin\theta_s)^{\mu'-1}(\cos\theta_s)^{\nu'-1}, \tag{66}$$

where

$$\mu' = 2(s-1-\bar{\bar{k}}_{s-1}-\tilde{\tilde{k}}_{s-1}-k''_{s-1}) + \bar{k}_{s-1} + \bar{k}'_{s-1} + \tilde{k}_{s-1} + \tilde{k}'_{s-1} + k_{s-1} + k'_{s-1}, \tag{67}$$

$$\nu' = 2 + \bar{k}_s + \tilde{k}_s + k_s + \bar{k}_{s-2} + \tilde{k}_{s-2} + k_{s-2} - \bar{k}_{s-1} - \bar{k}'_{s-1} - \tilde{k}_{s-1} - \tilde{k}'_{s-1} - k_{s-1} - k'_{s-1}, \tag{68}$$

and Q_2 is defined by (61) if $F^{(2)}_{s-1}$ is substituted instead of $F^{(1)}_{s-1}$. The parameter μ' of (67) is positive because of a summation condition in (66). According to (65) the parameter ν' of (68) is positive. The relations (55), (57), (60), and (64) express the sums

$$\sum_\gamma \left\langle \begin{matrix} \bar{m}_n & \tilde{m}_n & m_n \\ \bar{m}_{n-1} & \tilde{m}_{n-1} & m_{n-1} \\ \vdots & \vdots & \vdots \end{matrix} \middle| \gamma \right\rangle^* \left\langle \begin{matrix} \bar{m}_n & \tilde{m}_n & m_n \\ \bar{m}'_{n-1} & \tilde{m}'_{n-1} & m'_{n-1} \\ \vdots & \vdots & \vdots \end{matrix} \middle| \gamma \right\rangle \tag{69}$$

through the same sums but written for the subgroup $\mathrm{U}(n-1)$, and therefore, define completely these sums.

Separation of different summands in (69) is a complicated problem and we do not concern it here. This separation is not unique. If the tensor product of the representations $T^{\bar{m}_n}$ and $T^{\tilde{m}_n}$ contains the representation T^{m_n} with unit multiplicity, then (69) is reduced to one summand, the latter be defined completely. Separation of the free CGC in this case is fulfilled in the same manner as in Ref. 30 (for the case of the Lorentz group). We omit therefore the discussion of this procedure.

Thus, we have obtained all those CGCs of the group $\mathrm{U}(n)$, for which the resulting representation is contained in the tensor product with a unit multiplicity.

X. ON THE MATRIX ELEMENTS OF UNITARY AND FINITE DIMENSIONAL REPRESENTATIONS OF THE GROUP SO_0 $(n,1)$

For the groups $\mathrm{SO}_0(n,1)$ $K \approx \mathrm{SO}(n)$, $M \approx \mathrm{SO}(n-1)$, $A_+ = \exp(a_+)$, where a_+ is the one-dimensional subalgebra of $\mathrm{so}(n,1)$ [the Lie algebra of $\mathrm{SO}_0(n,1)$] with the basis element $e = E_{n,n+1} + E_{n+1,n}$. Hence an elementary representation of $\mathrm{SO}_0(n,1)$ is defined by an irreducible unitary representation δ of $\mathrm{SO}(n-1)$ and by a linear form Λ on a_+. The representation δ is given by the set of numbers $m_1,m_2,\ldots,m_{[(n-1)/2]}$ which are all integers or all half-integers (a half-integer means half of an odd integer) such that $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_{p-1} \geqslant 0$ for $\mathrm{SO}(2p-1)$ and $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_{p-1} \geqslant |m_p|$ for $\mathrm{SO}(2p)$. We shall use the numbers $l_1,l_2,\ldots,l_{[(n-2)/2]}$ where $l_i = m_i + [n/2] - i$. The linear form Λ is given by the number $\lambda = \Lambda(e)$. We shall use the number $c = \lambda - [(n-1)/2]$ instead of the number λ. The elementary representation $\pi_{\delta,\Lambda}$ will be denoted as $\pi(l_1,l_2,\ldots,l_{[(n-1)/2]},c) \equiv \pi(\mathrm{l},c)$.

There is only one nonidentical element w in the Weyl group W of the pair $[\mathrm{so}(n,1),a_+]$, and

$$w\pi(\mathrm{l},c) = \pi(\tilde{\mathrm{l}},-c) \equiv \pi(l_1,l_2,\ldots,l_{p-1},-l_p,-c) \tag{70}$$

in the case of $\mathrm{SO}_0(2p+1,1)$, and

$$w\pi(\mathrm{l},c) = \pi(\mathrm{l},1-c) \tag{71}$$

in the case of $\mathrm{SO}_0(2p,1)$.

We choose a basis in the space of the representation $\pi(\mathrm{l},c)$ in such a way that it coincides with the union of the orthonormal Gel'fand–Zetlin bases of subspaces of UIRs of $\mathrm{SO}(n)$. Hence, infinitesimal generators of $\mathrm{SO}(n)$ are given in this basis by well-known Gel'fand–Zetlin relations. We call this basis the canonical basis. Basis elements are denoted by $|m_n,\alpha\rangle$, where

$m_n = (m_{1n}, m_{2n}, \ldots, m_{[n/2],n})$ are the numbers, which define UIR of SO(n), and α is the Gel'fand–Zetlin scheme. It is shown in Ref. 18 that the canonical basis can be chosen in such a way that the infinitesimal operator $E_{n+1,n}$ of the representation $\pi(l,c)$ is given by

$$E_{2p+2,2p+1}|m_{2p+1},\alpha\rangle = \sum_{s=1}^{p}(c + l_{s,2p+1})\omega_s(l;m_{2p+1},\alpha)|m_{2p+1}^{+s},\alpha\rangle - \sum_{s=1}^{p}(c - l_{s,2p+1} + 1)\omega_s(l;m_{2p+1}^{-s},\alpha)$$
$$\times|m_{2p+1}^{-s},\alpha\rangle + \left[c\prod_{r=1}^{p} l_r l_{r,2p} \Big/ \prod_{r=1}^{p}(l_{r,2p+1} - 1)l_{r,2p+1}\right]|m_{2p+1},\alpha\rangle \tag{72}$$

in the case of $SO_0(2p,1)$, and by

$$E_{2p+1,2p}|m_{2p},\alpha\rangle = \sum_{s=1}^{p}(c + l_{s,2p})\omega_s(l;m_{2p},\alpha)|m_{2p}^{+s},\alpha\rangle - \sum_{s=1}^{p}(c - l_{s,2p})\omega_s(l;m_{2p}^{-s},\alpha)|m_{2p}^{-s},\alpha\rangle \tag{73}$$

in the case of $SO_0(2p,1)$. In (72) and (73)

$$\omega_s(l;m_{2p+1},\alpha) = \left|\frac{\Pi_{r=1}^{p}(l_r^2 - l_{s,2p+1}^2)(l_{r,2p}^2 - l_{s,2p+1}^2)}{l_{s,2p+1}^2(4l_{s,2p+1}^2 - 1)\Pi_{r=1,r\neq s}^{p}(l_{r,2p+1}^2 - l_{s,2p+1}^2)[(l_{r,2p+1} - 1)^2 - l_{s,2p+1}^2]}\right|^{1/2}, \tag{74}$$

$$\omega_s(l;m_{2p},\alpha) = \left|\frac{\Pi_{r=1}^{p-1}(l_{r,2p-1} + l_{s,2p})(l_{r,2p-1} - l_{s,2p} - 1)(l_r - l_{s,2p})(l_r - l_{s,2p} - 1)}{\Pi_{r=1,r\neq s}^{p}(l_{r,2p}^2 - l_{s,2p}^2)[l_{r,2p}^2 - (l_{s,2p} - 1)^2]}\right|^{1/2}, \tag{75}$$

and $l_{in} = m_{in} - i + [(n+1)/2]$.

The representation $\pi(l,c)$ of $SO_0(2p+1,1)$ is completely (or infinitesimally) irreducible iff neither c nor $-c$ is compared with $l_1, l_2, \ldots, l_p$, or one of the numbers c, $-c$ coincides with one of the numbers $l_1, l_2, \ldots, l_p$, or $\pm c < |l_p|$. The representation $\pi(l,c)$ of $SO_0(2p,1)$ is completely (or infinitesimally) irreducible iff c and $1-c$ are not compared with $l_1, l_2, \ldots, l_{2p-1}$, or one of the numbers c, $1-c$ coincides with one of the numbers $l_1, l_{2p}, \ldots, l_{p-1}$. Here "$c$ is compared with $l_1, l_2, \ldots, l_{[(n-1)/2]}$" means that c is an integer for integral l_i or c is a half-integer for half-integral l_i.

The principal unitary series of $SO_0(2p+1,1)$ consists of the representations $\pi(l,c)$, for which $c = i\rho$, where ρ is real number, and that of $SO_0(2p,1)$ consists of $\pi(l,c)$, for which $c = \frac{1}{2} + i\rho$, ρ be real number. The elementary representation $\pi(l,c)$ contains a finite dimensional representation iff c is compared with $l_1, l_2, \ldots, l_{[(n-1)/2]}$, and $c > l_1$.

As in the case of $U(n,1)$, a completely irreducible representation of $SO_0(n,1)$ is uniquely defined (up to infinitesimal equivalence) by its infinitesimal character and the set of its UIRs of SO(n). An infinitesimal character is given by the numbers (l,c), which define some elementary representation of $SO_0(n,1)$ with the same infinitesimal character.

Besides the irreducible elementary and finite dimensional representations, the group $SO_0(2p+1,1)$ has completely irreducible representations[20] $D^j(l,c), j = 1,2,\ldots,p-1$, where c is compared with $l_1, l_2, \ldots, l_p$ and $l_j > c > |l_{j+1}|$. $D_j(l,c)$ contains (with unit multiplicity) those and only those UIRs of $SO(2p+1)$, which are contained in the representation $\pi(l,c)$ and, in addition, satisfy the condition $l_{j+1,2p+1} \leqslant c$.

Besides the irreducible elementary and finite dimensional representations, the group $SO_0(2p,1)$ has completely irreducible representations $D^j(l,c), j = 1,2,\ldots,p-1$ where c is compared with $l_1, l_2, \ldots, l_{p-1}$ and $l_j > c > l_{j+1}$ for $j = 1,2,\ldots,p-2$ and $l_{p-1} > c > 0$ for $j = p-1$, and completely irreducible representations $D^{\mp}(l,c)$, where $l_{p-1} > 1$ and c is compared with $l_1, l_2, \ldots, l_{p-1}$ and $l_{p-1} > c > 0$. The representations $D^j(l,c)$, $D^{\mp}(l,c)$ contain those and only those UIRs of $SO(2p)$ which are contained in the representation $\pi(l,c)$ and, in addition, satisfy the condition $|l_{j+1,2p}| < c$ for $D^j(l,c)$, the condition $l_{p,2p} \geqslant c$ for $D^+(l,c)$, the condition $-l_{p,2p} \geqslant c$ for $D^-(l,c)$.

The following irreducible representations of $SO_0(2p+1,1)$ are unitarizable:

(1) $\pi(l,i\rho)$, if ρ is real;

(2) $\pi(l,\delta)$, where $0 < \sigma < s$ and s is an integer such that $l_{p-r+1} = r - 1$ under $r = 1,2,\ldots,s$;

(3) $D^j(l,p-j), j = 1,2,\ldots,p-1$, for which $l_{p-r+1} = r-1$, $r = 1,2,\ldots,p-j$. The following irreducible representations of $SO_0(2p,1)$ are unitarizable:

(1) $\pi(l,\frac{1}{2} + i\rho)$ if ρ is real;

(2) $\pi(l,\sigma + \frac{1}{2})$, where $0 < \sigma < s + \frac{1}{2}$ and s is an integer such that $l_{p-r} = r$, $r = 1,2,\ldots,s$; and $s = 0$, if $l_{p-1} > 1$;

(3) $D^j(l,p-j), j = 1,2,\ldots,p-1$, for which under $j \neq p-1$ $l_{p-r} = r$, $r = 1,2,\ldots,p-j-1$;

(4) $D^{\mp}(l,c)$.

All these representations are unitary if in their spaces orthonormal basis are chosen in such a way that the infinitesimal operator $E_{n+1,n}$ is given by

$$E_{2p+2,2p+1}|m_{2p+1},\alpha\rangle_s$$
$$= \sum_{i=1}^{p}(c^2 - l_{i,2p+1}^2)^{1/2}\omega_i(l;m_{2p+1},\alpha)|m_{2p+1}^{+i},\alpha\rangle_s$$
$$- \sum_{i=1}^{p}[c^2 - (l_{i,2p+1} - 1)^2]^{1/2}\omega_i(l;m_{2p+1}^{-i},\alpha)|m_{2p+1}^{-i},\alpha\rangle_s$$
$$+ c\prod_{r=1}^{p}[l_r l_{r,2p}/l_{r,2p+1}(l_{r,2p+1} - 1)]|m_{2p+1},\alpha\rangle_s \tag{76}$$

for the case of $SO_0(2p+1,1)$, and by

$$E_{2p+1,2p}|m_{2p},\alpha\rangle_s$$

$$= \sum_{i=1}^{p} [(c-\tfrac{1}{2})^2 - (l_{i,2p}+\tfrac{1}{2})^2]^{1/2}\omega_i(l,m_{2p},\alpha)|m_{2p}^{+i},\alpha\rangle_s$$

$$- \sum_{i=1}^{p} [(c-\tfrac{1}{2})^2 - (l_{i,2p}-\tfrac{1}{2})^2]^{1/2}\omega_i(l;m_{2p}^{-i},\alpha)|m_{2p}^{-i},\alpha\rangle_s \quad (77)$$

for the case of $SO_0(2p,1)$. In (76) and (77) ω_i are defined by (74) and (75), respectively.

The principal unitary series representations are unitary in the Hilbert space $L_\delta^2(K,H_\delta)$. For the unitarization of other unitarizable representations it is necessary to redefine a scalar product in the spaces of the elementary representations. As in the case of the group $U(n,1)$ the operator A of transition to unitarized representations of $SO(n,1)$ is associated with the intertwining operator corresponding to the nonidentical element w of the Weyl group W [see (70) and (71)]. As in the case of the group $U(n,1)$ intertwining operators of the elementary representations of $SO_0(n,1)$ are diagonal in the canonical basis, and their diagonal MEs do not depend on α.

If the representation $\pi(l,c)$ of $SO_0(2p+1,1)$ is irreducible, then the intertwining operator $\Pi^{l,c}_{l,-c,\tilde{l}} = (l_1,l_2,\ldots,l_{p-1},-l_p)$, is defined by the diagonal MEs

$$\lambda_{m_{2p-1}} = a(\tilde{l},-c,m_{2p+1})/a(l,c,m_{2p+1}), \quad (78)$$

where

$$a(l,c,m_{2p+1}) = \prod_{j=1}^{p}\prod_{\tau=l_j+1}^{l_{j,2p}-1}(c+\tau).$$

If the representation $\pi(l,c)$ of $SO_0(2p,1)$ is irreducible, then the intertwining operator $\Pi^{l,c}_{l,1-c}$ is defined by the numbers

$$\lambda_{m_{2p}} = a(l,1-c,m_{2p})/a(l,c,m_{2p}), \quad (79)$$

where

$$a(l,c,m_{2p}) = \prod_{j=1}^{p-1}\prod_{\tau=l_j}^{l_{j,2p}-1}(c+\tau)\prod_{\sigma=-l_{p,2p-1}+1}^{l_{p,2p}-1}(c+\sigma).$$

At fixed $l_1,l_2,\ldots,l_{[(n-1)/2]}$, $\Pi^{l,c}_{l,-c}$ in the case of $SO_0(2p+1,1)$ and $\Pi^{l,c}_{l,1-c}$ in the case of $SO_0(2p,1)$ are operator-valued analytic functions of c, which can be meromorphically continued into the whole of the complex plane. The continued operator function will be denoted by $\tilde{\Pi}^{l,c}_{l,-c}$ and $\tilde{\Pi}^{l,c}_{l,1-c}$ respectively. The function $\tilde{\Pi}^{l,c}_{l,-c}$ is regular [and is an intertwining operator for the representations $\pi(l,c)$ and $\pi(\tilde{l},-c)$] at all points of complex plane except for the points c which are compared with $l_1,l_2,\ldots,l_p$ and, in addition, satisfy one of the conditions $-c>l_1$; $l_j>-c>l_{j+1}$, $j=1,2,\ldots,p-2$; $l_{p-1}>-c>|l_p|$. At these points $\tilde{\Pi}^{l,c}_{l,-c}$ has simple poles, and corresponding residues are intertwining operators for the representations on $\pi(l,c)$ and $\pi(\tilde{l},-c)$ of $SO_0(2p+1,1)$. These intertwining operators are defined by the numbers $\lambda_{m_{2p}}$, which are residues of the numbers (78), as functions of c, at corresponding points. All the considered intertwining operators of $SO_0(2p+1,1)$ will be denoted by $\Pi^{l,c}_{l,-c}$.

The function $\tilde{\Pi}^{l,c}_{l,1-c}$ is regular [and is an intertwining operator for the representations $\pi(l,c)$ and $\pi(l,1-c)$ of $SO_0(2p,1)$] at all points of the complex plane except for the points c, which are compared with $l_1,l_2,\ldots,l_{p-1}$, and, in addition, satisfy one of the conditions $1-c>l_1$; $l_i>1-c>l_{i+1}$, $i=1,2,\ldots,p-2$; $l_{p-1}>c>\frac{1}{2}$. At the points indicated $\tilde{\Pi}^{l,c}_{l,1-c}$ has simple poles, and corresponding residues are intertwining operators for the representations $\pi(l,c)$ and $\pi(l,1-c)$ of $SO_0(2p,1)$. These intertwining operators are defined by the residues of the numbers (79) as functions of c, at corresponding points. All the considered intertwining operators of $SO_0(2p,1)$ will be denoted by $\Pi^{l,c}_{l,1-c}$.

Unitarizable representations are unitary if new orthonormal bases are chosen such that the noncompact infinitesimal operator $E_{n+1,n}$ is given by (76) for $SO_0(2p+1,1)$ and by (77) for $SO_0(2p,1)$. Such a basis will be called the standard basis. A basis of a finite dimensional representation of $SO_0(n,1)$ in which MEs can be continued to MEs of UIRs of $SO_0(n+1)$ will be called the standard basis too.

Theorem 6: Let $\Pi^{l,c}_{l',c'}$ be an intertwining operator for $\pi(l,c)$ and $\pi(l',c')$, where $(l',c') = (\tilde{l},-c)$ for $SO_0(2p+1,1)$ and $(l',c') = (l,1-c)$ for $SO_0(2p,1)$. Let H be the closure of the range of $\Pi^{l,c}_{l',c'}$ [H coincides with the space of the representation $\pi(l',c')$ if $\Pi^{l,c}_{l',c'}$ is reversible]. If the restriction of $\pi(l',c')$ onto H is a unitarizable or a finite dimensional representation, then the operator A of a transition from the canonical basis $|m_n,\alpha\rangle$ to the standard basis $|m_n,\alpha\rangle_s$ of H is given by

$$|m_n,\alpha\rangle_s = A\,|m_n,\alpha\rangle$$
$$= (\sqrt{-1})^a\langle m_n,\alpha|\Pi^{l,c}_{l',c'}|m_n,\alpha\rangle^{1/2}|m_n,\alpha\rangle, \quad (80)$$

where $a = \Sigma_{s=1}^{p} l_{s,n}$.

The proof of the theorem is carried out directly by evaluation of MEs of the operator A, the relations (72), (73), (76), and (77) being used.

Introduce the notation

$$\mu^{l,c}_{m_n} = (\sqrt{-1})^a\langle m_n,\alpha|\Pi^{l,c}_{l',c'}|m_n,\alpha\rangle^{1/2}.$$

To obtain MEs of unitary representation from MEs of the elementary representations it is sufficient to know the ratios $\mu_{m'_n}/\mu_{m_n}$. Theorem 6 implies that

$$\frac{\mu^{l,c}_{m_{2p-1}}}{\mu^{l,c}_{m'_{2p-1}}} = \prod_{s=1}^{p}\left(\frac{\Gamma(c+l'_{s,2p+1})\Gamma(c-l'_{s,2p+1}+1)}{\Gamma(c+l_{s,2p+1})\Gamma(c-l_{s,2p+1}+1)}\right)^{1/2} \quad (81)$$

for the group $SO_0(2p+1,1)$ and

$$\frac{\mu^{l,c}_{m_{2p}}}{\mu^{l,c}_{m'_{2p}}} = \prod_{s=1}^{p}\left(\frac{\Gamma(c-l'_{s,2p})\Gamma(c+l'_{s,2p})}{\Gamma(c-l_{s,2p})\Gamma(c+l_{s,2p})}\right)^{1/2} \quad (82)$$

for the group $SO_0(2p,1)$. The ratios $\Gamma(-z_1)/\Gamma(-z_2)$ with nonnegative integers z_1 and z_2 can appear in (81) and (82). In this case $\Gamma(-z_1)/\Gamma(-z_2)$ must be replaced by $(-1)^{z_1+z_2}\Gamma(1+z_2)/\Gamma(1+z_1)$.

According to Theorem 6, for unitarization of unitarizable representations of $SO_0(n,1)$ it is necessary to find the elementary representations which contain given unitarizable representations as subrepresentations. The composition series of the elementary representations of $SO_0(n,1)$ are de-

scribed in Ref. 20. According to Ref. 20, the unitarizable representation $D^j(l,p-j)$ of $SO_0(2p+1,1)$ is a subrepresentation of $\pi(\tilde{l}, -(p-j))$, $\tilde{l} = (l_1,l_2,...,l_{p-1}, -l_p)$. The finite dimensional representation $\Omega(c,l)$ of $SO_0(2p+1,1)$ with highest weight $m_{1,2p+2}, m_{2,2p+2},...,m_{p+1,2p+2}$, where $m_{1,2p+2} = c-p$, $m_{i,2p+2} = l_{i-1} - p + i - 1$, $i = 2,3,...,p+1$, is a subrepresentation of the representation $\pi(\tilde{l}, -c)$. The representations $D^j(l,p-j)$ and $D^{\pm}(l,c)$ of $SO_0(2p,1)$ are subrepresentations of $\pi(l,1-p+j)$ and $\pi(l,c)$, respectively. The finite dimensional representation $\Omega(c,l)$ of $SO_0(2p,1)$ with the highest weight $m_{1,2p+1}$, $m_{2,2p+1},...,m_{p,2p+1}$, where $m_{1,2p+1} = c-p$, $m_{i,2p+1} = l_{i-1} - p + i - 1$, $i = 2,3,...,p$, is a subrepresentation of the representation $\pi(l,1-c)$. MEs of the operator T of unitarizable and finite dimensional representation in the standard basis are given as

$$ {}_s\langle m_n,\alpha|T|m_n',\alpha'\rangle_s = \frac{\mu_{m_n''}^{w(l',c')}}{\mu_{m_n}^{w(l',c')}} \langle m_n,\alpha|T^{l',c'}|m_n',\alpha'\rangle, \tag{83} $$

where in the right-hand side there are MEs (in the canonical basis) of the same operator but taken for that elementary representation $\pi(l',c')$ (defined above) which contains given unitarizable or finite dimensional representation as subrepresentation. For the representations $\pi(l,\sigma)$ of the classes (2) of the classification of the unitary representations of $SO_0(n,1)$, MEs in the standard basis are given by the relation (83) with $(l',c') = (l,\sigma)$. MEs in the right-hand side of (83) are the analytic continuation in c' of MEs of the principal unitary series representations in the canonical basis of $L^2_\delta(K,H_\delta)$. A transition from MEs of finite dimensional representations of $SO_0(n,1)$ in the standard basis to MEs of UIRs of $SO(n+1)$ is carried out in the same way as for the groups $U(n,1)$ and $U(n+1)$ (see Sec. IX).

Unfortunately, MEs are found not for all the principal unitary series representations. Only MEs of spherical representations of the principal unitary series of $SO_0(n,1)$ are known (cf. Ref. 31). In Ref. 32 the relation between MEs of the principal unitary series representations of $SO_0(n,1)$ and MEs of UIRs of $SO(n+1)$ is given. In Ref. 33 the relation between MEs of the spherical representations of the principal unitary series and of the complementary series was considered. It was also pointed in Ref. 33 that such a relation concerns intertwining operators of spherical representations.

XI. ANALYTIC PROPERTIES OF CLEBSCH–GORDAN COEFFICIENTS

The semisimple (or reductive) Lie algebra $\mathfrak{g}$ decomposes into the direct sum $\mathfrak{g} = \mathfrak{j} + \mathfrak{p}$, where $\mathfrak{j}$ is a maximal compact Lie subalgebra in $\mathfrak{g}$. Since $[\mathfrak{j},\mathfrak{p}] \subset \mathfrak{p}$, $\mathfrak{p}$ is a space of the representation of $\mathfrak{j}$. Denote this representation by D. Let p_i be basis elements of $\mathfrak{p}$. The operators P_i of the elementary representation $d\pi_{\delta,\Lambda}$ of $\mathfrak{g}$, which correspond to the elements p_i, are the components of tensor operator transforming according to the representation D of $\mathfrak{j}$. Therefore, the Wigner–Eckart theorem can be applied to P_i. In the space of the elementary representation $d\pi_{\delta,\Lambda}$ a basis is chosen, which consists of bases of irreducible representations of $\mathfrak{j}$. Denote these basis elements by $|\Lambda,\delta,\lambda,n\rangle$, where λ are the numbers which define UIRs of $\mathfrak{j}$, and n labels basis elements in UIR of $\mathfrak{j}$. We assume that the basis elements $|\Lambda,\delta,\lambda,n\rangle$ do not depend on Λ. The existence of such bases follows from the definition of a space of an elementary representation. Then, according to the Wigner–Eckart theorem,

$$ P_i|\Lambda,\delta,\lambda,n\rangle = \sum_{j=1}^{r} C_{\lambda\lambda_j}(\Lambda,\delta)\langle\lambda,n;D,i|\lambda_j,n_j\rangle|\Lambda,\delta,\lambda_j n_j\rangle, \tag{84} $$

where r is a dimension of the space $\mathfrak{p}$, $C_{\lambda\lambda_j}(\Lambda,\delta)$ are the numbers not depending on i, n_j, n, and $\langle\cdots|\cdots\rangle$ are CGCs of the subalgebra $\mathfrak{j}$. Dependence on Λ is completely contained in the coefficients $C_{\lambda\lambda_j}(\Lambda,\delta)$. For the elementary representations the coefficients $C_{\lambda\lambda_j}(\Lambda,\delta)$ depend linearly[34] on $c_1,c_2,...,c_l$ (the coordinates of Λ). This fact will be used below.

As usual, CGCs of the decomposition of the tensor product of the elementary representations $\pi_{\delta',\Lambda'}$ and $\pi_{\delta'',\Lambda''}$ into elemenary representations $\pi_{\delta,\Lambda}$ are introduced according to

$$ |\Lambda,\delta,\lambda.n\rangle = \sum_{\substack{\lambda',n'\\ \lambda''n''}} \langle\Lambda',\delta',\lambda',n';\Lambda'',\delta'',\lambda'',n''|\Lambda,\delta,\lambda,n\rangle \times|\Lambda',\delta',\lambda',n'\rangle\otimes|\Lambda'',\delta'',\lambda'',n''\rangle. \tag{85} $$

As a consequence of (85), it is easy to obtain the following relation for the infinitesimal operator I:

$$ \begin{aligned} &\sum_{\lambda_1',n_1'} \langle\Lambda',\delta',\lambda_1',n_1';\Lambda'',\delta'',\lambda'',n''|\Lambda.\delta,\lambda,n\rangle \\ &\times D^{\Lambda',\delta'}_{(\lambda_1',n_1')(\lambda',n')}(\mathrm{I}) + \sum_{\lambda_1'',n_1''} \langle\Lambda',\delta',\lambda',n';\Lambda'', \\ &\times\delta'',\lambda_1'',n_1''|\Lambda,\delta,\lambda,n\rangle D^{\Lambda'',\delta''}_{(\lambda_1'',n_1'')(\lambda'',n'')}(\mathrm{I}) \\ &= \sum_{\lambda_1,n_1} D^{\Lambda,\delta}_{(\lambda,n)(\lambda_1,n_1)}(\mathrm{I}) \\ &\times\langle\Lambda',\delta',\lambda',n';\Lambda'',\delta'',\lambda'',n''|\Lambda,\delta,\lambda_1,n_1\rangle. \end{aligned} \tag{86} $$

In fact we may write the relation (85) for the tensor product of the principal unitary series representations. This tensor product is decomposed into a direct integral of the principal unitary series representations. In order to study analytic properties of CGCs, we write the relation (85) for any elementary representations and call the coefficients of (85) CGCs too. If the tensor product of the elementary representations $\pi_{\delta',\Lambda'}$ and $\pi_{\delta'',\Lambda''}$ does really contain in the decomposition the elementary representation $\pi_{\delta,\Lambda}$, then CGCs of this decomposition satisfy the relations (85) and (86). Moreover, if we consider in (85) and (86) only those basis vectors $|\Lambda',\delta',\lambda',n'\rangle$ and $|\Lambda'',\delta'',\lambda'',n''\rangle$ which belong to the spaces of subrepresentations π' and π'' of the elementary representatins $\pi_{\delta',\Lambda'}$ and $\pi_{\delta'',\Lambda''}$, respectively, and if the decomposition of the tensor product of subrepresentations π' and π'' contains the elementary representation $\pi_{\delta,\Lambda}$ or its subrepresentation, then CGCs of this decomposition satisfy the relations (85) and (86). Hence, studying the relations (85) and (86) for all elementary representations $\pi_{\delta,\Lambda}$, $\pi_{\delta',\Lambda'}$, $\pi_{\delta'',\Lambda''}$ we include in our investigation really existing CGCs of the group G.

The coefficients of the relations (85) and (86) for any elementary representations $\pi_{\delta,\Lambda}$, $\pi_{\delta',\Lambda'}$, $\pi_{\delta'',\Lambda''}$ are directly

related to the invariant under G threelinear forms on the spaces $dL^2_\delta(K,H_\delta)$, $dL^2_{\delta'}(K,H_{\delta'})$, $dL^2_{\delta''}(K,H_{\delta''})$. The invariant under Lorentz group threelinear forms in continuous basis were studied in Ref. 35.

Suppose that the tensor product of the representations $\pi_{\delta',\Lambda'}$ and $\pi_{\delta'',\Lambda''}$ is decomposed into a direct integral of the representations $\pi_{\delta,\Lambda}$. This integral is taken over some measure. This measure depends on Λ and δ, but it does not depend on λ and n. The relations (85) and (86) do not feel this measure. Hence, a solution of the system of equations (86) does not depend on this measure. In order to obtain CGCs of the decomposition of the tensor product $\pi_{\delta',\Lambda'}\otimes\pi_{\delta'',\Lambda''}$ into the direct integral, this solution must be multiplied by a function which depends on Λ and δ. Generally speaking this function (denote it by F) can be a generalized function. But it is not uniquely defined. Really, the direct integral of the representations $\pi_{\delta,\Lambda}$ over the measure $\mu(\Lambda,\delta)$ is unitarily equivalent to the direct integral over the measure $\mu'(\Lambda,\delta)$ if both these measures are absolutely continuous with respect to each other.[36] To different measures there correspond different functions F.

We neglect the function F and consider CGCs as usual numbers depending on Λ and δ. In fact the function F defines a normalization of CGCs. Hence we neglect a normalization condition, which depends on the measure. If we are interested in the invariant three-linear forms, the function F is of no importance because these forms are defined up to a constant.

The tensor product of the principal unitary series representations can contain the representation $\pi_{\delta,\Lambda}$ with some multiplicity. For this reason we rewrite the relation (85) as

$$|\Lambda,\delta,\lambda,n\rangle^\gamma = \sum_{\substack{\lambda',n'\\ \lambda'',n''}} \langle\Lambda',\delta',\lambda',n';\Lambda'',\delta'',\lambda'',n''|\Lambda,\delta,\lambda,n\rangle^\gamma \times|\Lambda',\delta',\lambda',n'\rangle\otimes|\Lambda'',\delta'',\lambda'',n''\rangle, \qquad (87)$$

where γ labels multiple representations. Since MEs of the elementary representations do not depend on γ, CGCs $\langle\Lambda',\delta',\lambda',n';\Lambda'',\delta'',\lambda'',n''|\Lambda,\delta,\lambda,n\rangle^\gamma$ with different γ satisfy the same relation (86). Let $\langle\Lambda',\delta',\lambda',n';\Lambda'',\delta'',\lambda'',n''|\Lambda,\delta,\lambda,n\rangle^\gamma$, $\gamma=1,2,...,m$ (m is the multiplicity of the representation $\pi_{\delta,\Lambda}$) be a set of independent solutions of the system of the relations (86). Acting upon this set of solutions by a nondegenerate $n\times n$ matrix we obtain a new system of solutions. Therefore, in order to obtain a fixed system of solutions it is necessary to fix some initial CGCs which define uniquely all CGCs with the same $\Lambda,\delta,\Lambda',\delta',\Lambda'',\delta''$. A number of initial CGCs depends on the multiplicity m of $\pi_{\delta,\Lambda}$ in the tensor product $\pi_{\delta',\Lambda'}\otimes\pi_{\delta'',\Lambda''}$.

Theorem 7: Let the multiplicity of any representation $\pi_{\delta,\Lambda}$ of the principal unitary series of G in the tensor product of representations $\pi_{\delta',\Lambda'}$ and $\pi_{\delta'',\Lambda''}$ of the principal unitary series with fixed δ', and δ'' equal 0 or m, where m is a fixed integer. Then CGC $\langle\Lambda',\delta',\lambda',n';\Lambda'',\delta'',\lambda'',n''|\Lambda,\delta,\lambda,n\rangle^\gamma$ of (87) for elementary representations, as functions of $\Lambda,\Lambda',\Lambda''$, is a ratio of polynomials of coordiantes $c_1,c_2,...,c_l,c_1',c_2',...,c_l',c_1'',c_2'',...,c_l''$ of the forms $\Lambda,\Lambda',\Lambda''$, multiplied by a function which depends on $\Lambda',\Lambda',\Lambda''$ and does not depend on $\lambda,\lambda',\lambda'',n,n',n''$.

Proof: Let us consider the relations (86) and (87) for CGCs which realize the decomposition of the tensor product of the principal unitary series representations $\pi_{\delta',\Lambda'}$ and $\pi_{\delta'',\Lambda''}$ into a direct integral of the principal unitary series representations. Fix δ,δ',δ''. Then CGCs are functions of $\Lambda,\Lambda',\Lambda''$. For all $\Lambda,\Lambda',\Lambda''$ we choose the same initial CGCs $\langle\Lambda',\delta',\lambda_0',n_0';\Lambda'',\delta'',\lambda_0'',n_0''|\Lambda,\delta,\lambda_0,n_0\rangle^\gamma$, which are independent on $\Lambda,\Lambda',\Lambda''$. The indices $\lambda_0,\lambda_0',\lambda_0''$ are chosen such that the corresponding UIRs of K are the lowest UIRs of K in the representations $\pi_{\delta,\Lambda},\pi_{\delta',\Lambda'},\pi_{\delta'',\Lambda''}$, respectively. According to the conditions of the theorem, a number of initial CGCs at fixed δ,δ',δ'' is the same for all $\Lambda,\Lambda',\Lambda''$. Consider the relations (86) for all linearly independent infinitesimal generators of the group G. All CGCs, as functions of $\Lambda,\Lambda',\Lambda''$, can be found successively from these relations. Namely, the relations (86) are considered for $\lambda=\lambda_0$, $\lambda'=\lambda_0'$, $\lambda''=\lambda_0''$. They define uniquely all those CGCs for which only one of the indices $\lambda,\lambda',\lambda''$ differs from the indices $\lambda_0,\lambda_0',\lambda_0''$ and equals to an index of the next UIR of K. Let $\bar\lambda,\bar\lambda',\bar\lambda''$ be one of the systems of indices for which CGCs have been found. Consider the relations (86) for $\lambda=\bar\lambda,\lambda'=\bar\lambda',\bar\lambda''=\bar\lambda''$. They define uniquely all CGCs for which only one of the indices $\lambda,\lambda',\lambda''$ differs from the indices $\bar\lambda,\bar\lambda',\bar\lambda''$ and equals to an index of the next UIR of K. Continuing this procedure, we find successively all CGCs as functions of the coordinates $c_1,c_2,...,c_l,c_1',c_2',...,c_l',c_1'',c_2'',...,c_l''$ of the forms $\Lambda,\Lambda',\Lambda''$. We do not give here the proof of the fact that at every step CGCs are found uniquely. This fact can be easily verified for every concrete semisimple Lie group. In general case, this proof can be carried out by the induction method if one considers the relations (86) for the basis elements P_i of the subspace $\mathfrak{p}$ of Lie algebra $\mathfrak{g}$ and takes into account the relation (84). The general proof of this fact is cumbersome and we omit it.

Let us return to the procedure of finding CGCs. At every step we have to solve a finite system of linear equations. As MEs of infinitesimal operators of the representation $\pi_{\delta,\Lambda}$ depend linearly on the coordinates of the form Λ, the solutions of this system are ratios of polynomials. Thus the theorem is proved for the principal unitary series representations. To prove the theorem for any elementary representations it is enough to continue analytically the relations (86) with CGCs of the principal unitary series representations to the space C^{3l} of complex parameters $c_1,c_2,...,c_l,c_1',c_2',...,c_l',c_1'',c_2'',...,c_l''$. Due to Theorem 1 CGCs for the elementary representations are an analytic continuation of CGCs for the principal unitary series representations. This completes the proof.

According to Theorem 2, some elementary representations of G contain finite dimensional representations of G as subrepresentations. This leads to the following Corollary of Theorem 7.

Corollary: If group G satisfies the conditions of Theorem 7, then CGCs of the elementary representations of G are an analytic continuation of the corresponding CGCs of finite dimensional representations of G in appropriate bases.

Since the elementary representations of G contain all irreducible representations of G, we know all CGCs of irre-

ducible representations (in particular, of the principal unitary series representations, of the complementary series representations, of the discrete series representations) if we know CGCs of the elementary representations. The so obtained CGCs for unitary representations are not always CGCs in the orthonormal basis in which the representations are unitary. A transition to CGCs in the orthornormal basis can be realized by means of the operator A introduced in Sec. VI. CGCs for the tensor product of finite dimensional representations of G, obtained from CGCs of the elementary representations, do not concide with CGCs for the tensor product of UIRs of the compact form G_k of the complexification of G. The transition is realized by means of operator A too.

Theorem 7 and its corollary give a possibility to obtain symmetry and recurrence relations for CGCs of infinite dimensional representations if those of finite dimensional representations are known. In fact, all symmetry and recurrence relations, not depending on the normalization function, remain valid if a correction with the help of the operator A is done (if necessary).

An analytic continuation of CGCs of finite dimensional representations to those of the principal unitary series representations for the case of the Lorentz group $SO_0(3,1)$ was done in Ref. 37. However, for a transition to the orthonormal basis of the space $L^2_\delta(K,H_\delta)$ in Ref. 37 it would be necessary to do a correction with the help of the operator A (i.e., of the numbers $\mu_{m_n}/\mu_{m'_n}$ from Sec. X).

Remark: After the manuscript was submitted for publication, we learned of Vilenkin's paper,[38] in which the formulas identical to (48) and (49) were obtained by means of techniques different from ours. We also note that some infinitesimal aspects of the relations between the representations of the groups SU(2,1) and SU(3), and also of SU(p,1) and SU($p+1$) were discussed by Biedenharn and coauthors.[39,40]

[1]J.J. de Swart, Rev. Mod. Phys. **35**, 916 (1963); A. Pais, Rev. Mod. Phys. **38**, 215 (1966); S. Oneda and E. Takasugi, "SU(4)-Algebraic Approach to New Resonances," Technical Report No. 76-139, University of Maryland, 1976.
[2]M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Massachusetts, 1964); B.G. Wybourne, *Classical Groups for Physicists* (Wiley, Interscience, New York, 1974); V. Vanagas, *Algebraic Methods in the Theory of Nucleous* (Mintis, Vilnius, 1972) (in Russian).
[3]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966); A.O. Barut, P. Budini, and C. Fronsdal, J. Math. Phys. **13**, 577 (1972).
[4]V.K. Dobrev, G. Mack, V.B. Petkova, S.G. Petrova, and I.T. Todorov, IAS preprint, Princeton, 1975; V.K. Dobrev, V.B. Petkova, S.G. Petrova, and I.T. Todorov, IAS preprint, Princeton, 1975.
[5]J.D. Louck, J. Math. Phys. **6**, 1786 (1965); R.C. Hwa and J. Nuyts, Phys. Rev. **145**, 1188 (1966); E.B. Aronson, I.A. Malkin, and V.I. Manko, Element. Tchast. i Atomn. Yadro **5**, 122 (1974).
[6]Harish-Chandra, Trans. Amer. Math. Soc. **75**, 185 (1953); D.P. Zhelobenko, *Harmonic Analysis on Complex Semisimple Lie Groups* (Nauka, Moscow, 1974) (in Russian).
[7]G. Warner, *Harmonic Analysis on Semisimple Lie Groups* (Springer, Berlin and New York, 1972), Vol. 1.
[8]B. Kostant, J. Func. Anal. **20**, 257 (1975); A.U. Klimyk, Lett. Math. Phys. **1**, 375 (1977).
[9]A.U. Klimyk, Rep. Math. Phys. **7**, 153 (1975).
[10]R.P. Martin, Trans. Amer. Math. Soc. **201**, 177 (1975); F. Williams *Lecture Notes in Mathematics* (Springer, Berlin and New York, 1973), Vol. 358.
[11]K.B. Wolf, J. Math. Phys. **13**, 1634 (1972).
[12]S.I. Alisauskas, A.A.-A. Jucys, and A.P. Jucys, J. Math. Phys. **13**, 1329 (1972); E. Chacon, M. Ciftan, and L.C. Biedenharn, J. Math. Phys. **13**, 577 (1972); L.A. Shelepin, Trudy Fizich. Inst. im. P.N. Lebedeva **70**, 3 (1973).
[13]Harish-Chandra, Trans. Amer. Math. Soc. **76**, 234 (1954); J. Lepowsky, *ibid.* **176**, 1 (1973).
[14]J. Lepowsky and N.R. Wallach, Trans. Amer. Math. Soc. **184**, 223 (1973).
[15]A.W. Knapp and E.M. Stein, Ann. Math. (N.Y.) **93**, 489 (1971).
[16]G. Schiffmann, Bull. Soc. Math. France **99**, 3 (1971); R. Kunze and E.M. Stein, Amer. J. Math. **89**, 385 (1967).
[17]Harish-Chandra, Trans. Amer. Math. Soc. **70**, 28 (1951).
[18]A.M. Gavrilik and A.U. Klimyk, preprint ITP-75-63E, Kiev, 1975.
[19]A.M. Gavrilik and A.U. Klimyk, preprint ITP-75-18E, Kiev, 1975.
[20]A.U. Klimyk and A.M. Gavrilik, preprint ITP-76-39E, Kiev, 1976.
[21]A.U. Klimyk and V.A. Shirokov, preprint ITP-74-143E, Kiev, 1974.
[22]H. Kraljevič, Glasnic Mathematički **8**, 23 (1973).
[23]D.F. Holland, J. Math. Phys. **10**, 1903 (1969).
[24]R. Raczka and J. Fischer, Commun. Math. Phys. **3**, 233 (1966).
[25]I.M. Gelfand, M.I. Graev, Izv. AN SSSR, Ser. Matem. **29** 1329 (1965).
[26]R. Goodman, Trans. Amer. Math. Soc. **143**, 55 (1969).
[27]J. Lepowsky and G.M. McCollum, Trans. Amer. Math. Soc. **176**, 45 (1973).
[28]D.P. Zhelobenko, *Compact Lie Groups and Their Representations* (Nauka, Moscow, 1970) (in Russian).
[29]I.S. Gradshtein and I.M. Ryzhik, *Tables of Integrals, Sums, Series, and Products* (Academic, New York, 1965).
[30]R.L. Anderson, R. Raczka, M.A. Rashid, and R. Winternitz, J. Math. Phys. **11**, 1050 (1970).
[31]N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Amer. Math. Soc., Providence, Rhode Island, 1968); C.P. Boyer, J. Math. Phys. **12**, 1599 (1971).
[32]T. Maekawa, J. Math. Phys. **16**, 2372 (1975).
[33]C.P. Boyer, J. Math. Phys. **14**, 609 (1973).
[34]E. Thieleker, "Quasisimple and Unitary Representations of Semisimple Lie Groups. I." preprint of the South Florida University.
[35]A.I. Oksak, Commun. Math. Phys. **29**, 189 (1973); Rep. Math. Phys. **7**, 9 (1975).
[36]J. Dixmier, *Les algebres d'operateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1969).
[37]S.I. Alisauskas, Lietuvos Fiz. Rinkinys **13**, 829 (1973).
[38]N.Ya. Vilenkin, Sb. Nauchnych Trudov Mosk. Gos. Zaochn. Ped. Inst. Moscow **39**, 77 (1974) (in Russian).
[39]L.C. Biedenharn, J. Nuyts, and N. Straumann, Ann. Inst. Henri Poincaré, **3A**, 13 (1965).
[40]L.C. Bidenharn, in *Cargèse Lectures in Theoretical Physics* (Gordon and Breach, London, 1967).

Evaluation of SU(6)⊃SU(3) ⊗ SU(2) Wigner coefficients[a)]

D. Strottman[b)]

Department of Physics, State University of New York, Stony Brook, New York 11794
(Received 15 May 1978)

A technique is described which allows the calculation of any SU(6) Wigner coefficient (isoscalar factor) from known one-body SU(6) coefficients. The technique is applied to calculate the SU(6) coefficients for $q^3 \times q^3$ leading to SU(3) singlet, octet, and decuplet states of six quarks. A sum rule of SU(3) $9-\lambda\mu$ symbols is given.

I. INTRODUCTION

The SU(6) symmetry group has for several years been applied to classifying hadrons into supermultiplets. In addition to the apparently successful classification schemes, it has in some cases been possible to calculate decay rates, electromagnetic properties and, of course, masses of the hadrons.[1] Recently, the SU(6) model has been extended by Melosh[2] to calculate matrix elements of currents between hadronic states labelled by irreducible representation of SU(6), yielding satisfactory results for certian axial vector coupling constants and the nucleon magnetic moments. This technique has been applied as well to pion and photon decay processes of hadrons.[3,4]

In all such applications of the SU(6) model, the calculation proceeds by applying the Wigner–Eckert theorem to the relevant matrix element resulting in a product of SU(6), SU(3), and SU(2) Wigner coefficients (the former two are often referred to as isoscalar factors) and a reduced matrix element. The viability of such a procedure is contingent on the availability of the relevant SU(6) Wigner coefficient, the SU(3) and SU(2) coefficients being always calculable[5] irregardless of the complexity of the respresentation. The SU(6) coefficients have not, however, been calculated save for some special cases.[6]

In the usual quark model, the low-lying hadrons are constructed from the mimimum number of quarks possible, viz., $q\bar{q}$ for mesons and q^3 for hadrons. The possibility that some hadrons are rather states with an additional number of quarks has been investigated by Jaffe[7] using a variant of the MIT bag model[8] for the case of mesons constructed from the configuration $\bar{q}^2 q^2$. In Jaffe's version of the bag model the hadron states may be usefully labeled by the irreducible representations of SU(6), although in general SU(6) is not a symmetry of the Hamiltonian. [The SU(6) group is obtained by the omission of the much more massive charmed quark.] Jaffe concluded that certain mesons may be manifestations of such a configuration, these mesons having previously been incorrectly assumed to be P-wave quark antiquark states. An examination of the $q^4\bar{q}$ configuration was not possible for lack of the requisite SU(6) Wigner coefficients.

Recently, the SU(6)⊃SU(3) ⊗ SU(2) Wigner coefficients were calculated[9] for the case of a single quark coupling to a many quark state. It is the purpose of this paper to show how more complicated SU(6) Wigner coefficients may be calculated with the help of the known coefficients.[9] In particular all coefficients necessary for coupling an arbitrary number of quarks to an arbitrary number of quarks or antiquarks may be calculated. Similarly, any coefficient necessary in applying the Melosh transform is also calculable. In Sec. II notation will be introduced and in Sec III the method of calculation will be explained.

II. NOTATION

A ket of n quarks or antiquarks will be written as

$$\left|\begin{matrix}[f]=[f_1f_2f_3f_4f_5f_6] \\ (\lambda\mu)\quad S\quad \omega \\ I\quad I_3\quad Y\end{matrix}\right\rangle, \quad \lambda=g_1-g_2, \quad \mu=g_2-g_3,$$

where $[f]$, $(\lambda\mu)$, and S denote the irreducible representations of SU(6), SU(3), and SU(2), respectively. The physically interesting subgroup SU(6)⊃SU(3) ⊗ SU(2) is a noncanonical chain, implying that the group operators do not in general provide a sufficient number of labeling operators to uniquely specify the ket. The quantity ω is a multiplicity label which serves to distinguish otherwise identical states. The labels I, I_3, and Y are the isospin, isospin third component, and hypercharge.

In this paper the representations of SU(6) and SU(3) will be labeled by the Young tableaux $[f]$ and $[g]$, respectively. In addition to the labels which appear explicitly in the ket, there must also occur further SU(3) group labels. If the SU(3) subgroup of SU(6) is assumed to SU(3) color, then the additional SU(3) flavor group must also be present. The requirement of antisymmetry immediately imposes the condition that the SU(3) flavor representation be conjugate to the SU(6) representation. Since the SU(3) flavor representation is fixed once the SU(6) representation is specified, and since no operator which appears in the procedure to be outlined below depends on the SU(3) flavor representation, the labels will not be included in the basis states, but are to be implicitly understood.

For those who indulge in the vulgar habit of labeling the irreducible representations of the unitary group multiplets with their dimensionality, the dimensions for SU(6) and SU(3) representations are

[a)]Partially supported by USDOE Contract No. EY-76-S-02-3001.
[b)]Present address: Theoretical Division, Los Alamos Scientific Laboratory, Los Alamos, NM 87545.

 0022-2488/79/081643-05$01.00

TABLE I. Values of the Casimir operator of SU(6) for the allowed states of a two quark system.

$[f_{23}]$	$C^{(2)}([f_{23}])$	$(\lambda_{23}\mu_{23})S_{23}$
[2]	$\frac{40}{3}$	(20),1 (01)0
[11]	$\frac{28}{3}$	(20),1, (01)1

$$g([f]) = \prod_{i<j}^{6} \left(\frac{f_i - f_j + j - i}{j - i}\right), \tag{1}$$

$$g((\lambda\mu)) = \tfrac{1}{2}(\lambda+1)(\mu+1)(\lambda+\mu+2).$$

The expectation values of the Casimir invariants are

$$\begin{aligned} \mathscr{C}_6^{(2)} &= \sum_{\alpha}^{6} f_\alpha^2 - \frac{1}{6}\left(\sum f_\alpha\right)^2 + 5(f_1 - f_6) + 3(f_2 - f_5) \\ &\quad + f_3 - f_4, \\ \mathscr{C}_3^{(2)} &= \tfrac{2}{3}(\lambda^2 + \mu^2 + \lambda\mu + 3\lambda + 3\mu). \end{aligned} \tag{2}$$

III. PRODUCT STATE AND SU(6)⊃SU(3)⊗SU(2) WIGNER COEFFICIENTS

The SU(6)⊃SU(3)⊗SU(2) Wigner coefficients for which a single quark couples to a state of several quarks will be referred to henceforth as one-body Wigner coefficients. In this section it will be shown how to calculate the SU(6) Wigner coefficients appropriate for coupling several quark states to several quark states and the technique will then be generalized in a trivial manner to allow the inclusion of antiquark states.

Consider a basis state in which configurations of n_1 and n_2 quarks are separately members of irreducible representations of SU(6), SU(3), and SU(2). The product state may immediately be coupled to a representation of SU(3) and SU(2) through the use of the appropriate SU(3) and SU(2) Wigner coefficients,

$$\begin{aligned} &\left| q^{n_1} \begin{matrix}[f_1]\\(\lambda_1\mu_1)\end{matrix} S_1\omega_1 \otimes q^{n_2} \begin{matrix}[f_2]\\(\lambda_2\mu_2)\end{matrix} S_2\omega_2;\ (\lambda\mu)S\omega\rho \quad II_3Y \right\rangle \\ &= \sum_{\substack{Y_1I_1\\Y_2I_2}} \left\langle \begin{matrix}(\lambda_1\mu_1) & (\lambda_2\mu_2)\\ I_1Y_1 & I_2Y_2\end{matrix} \middle| \begin{matrix}(\lambda\mu)\\IY\end{matrix} \right\rangle_\rho \left[\left| q^{n_1} \begin{matrix}[f_1]\\(\lambda_1\mu_1)S_1\omega_1\end{matrix} I_1\ Y_1 \right\rangle \left| q^{n_2} \begin{matrix}[f_2]\\(\lambda_2\mu_2)S_2\omega_2\end{matrix} I_2\ Y_2 \right\rangle_{I_3M}^{IS} \right]. \end{aligned} \tag{3}$$

In Eq. (3) ρ is the outer multiplicity and the square bracket indicates the two kets are coupled to total isospin I and S with the indicated z projections. The correct linear combination of the basis states to produce a product state which is labeled by SU(6) may be found by diagonalizing the Casimir operator of SU(6),

$$\mathscr{C}_6^{(2)} = \sum_{\mu}^{35} \alpha_\mu \alpha_\mu, \quad \alpha_\mu \equiv \sum_{i}^{n_1+n_2} \alpha_\mu(i). \tag{4}$$

where the α_μ are the generators of SU(6).

The Casimir operator may be divided into three terms

$$\mathscr{C}_6^{(2)} = \mathscr{C}_6^{(2)}(n_1) + \mathscr{C}_6^{(2)}(n_2) + 2\sum_{\substack{i\in n_1\\ j\in n_2}} \sum_{\mu} \alpha_\mu(i)\alpha_\mu(j), \tag{5}$$

where the argument of $\mathscr{C}_6^{(2)}$ in the first two terms on the right-hand side indicates that the Casimir operator acts only on the group of n_1 or n_2 quarks. As the quark states in Eq. (3) are separately antisymmetric, any particular quark amongst the first n_1 quarks—say quark n_1—may be singled out from the sum over i, and similarly quark n_2 may be singled out from the sum over j, and the matrix element multiplied by n_1n_2,

$$\mathscr{C}_6^{(2)} = \mathscr{C}_6^{(2)}(n_1) + \mathscr{C}_6^{(2)}(n_2) + 2n_1n_2 \sum_{\mu} \alpha_\mu(n_1)\alpha_\mu(n_2). \tag{6}$$

The last term of Eq. (6) may be related to a two quark SU(6) Casimir operator,

$$2\sum_{\mu}\alpha_\mu(n_1)\alpha_\mu(n_2) = \sum_{\mu}[\alpha_\mu(n_1) + \alpha_\mu(n_2)]^2 - \sum_{\mu}\alpha_\mu(n_1)^2 - \sum_{\mu}\alpha_\mu(n_2)^2. \tag{7}$$

The expectation value of $\mathscr{C}_6^{(2)}$ may now be evaluated with the use of Eqs. (6) and (7).

$$\left\langle q^{n_1} \begin{matrix}[f_1]\\(\lambda_1\mu_1)S_1\omega_1\end{matrix} \otimes q^{n_2} \begin{matrix}[f_2]\\(\lambda_2\mu_2)S_2\omega_2\end{matrix} ;(\lambda\mu)S\omega\rho \middle| \mathscr{C}_6^{(2)} \middle| q^{n_1} \begin{matrix}[f_1]\\(\bar\lambda_1\bar\mu_1)\bar S_1\bar\omega_1\end{matrix} \otimes q^{n_2} \begin{matrix}[f_2]\\(\bar\lambda_2\bar\mu_2)\bar S_2\bar\omega_2\end{matrix} (\lambda\mu)S\omega\rho \right\rangle$$

$$= \delta_{1\bar{1}}\delta_{2\bar{2}}\{\langle \mathscr{C}_6^{(2)}([f_1])\rangle + \langle \mathscr{C}_6^{(2)}([f_2])\rangle\}$$
$$+ n_1 n_2 \left\langle \begin{matrix}[f_1]\\ (\lambda_1\mu_1)S_1\omega_1\end{matrix} \otimes \begin{matrix}[f_1]\\ (\lambda_2\mu_2)S_2\omega_2\end{matrix};(\lambda\mu)S\omega \middle| \sum_\mu [\alpha_\mu(n_1)+\alpha_\mu(n_2)]^2 \right.$$
$$\left. - \sum_\mu \alpha_\mu(n_1)^2 - \sum_\mu \alpha_\mu(n_2)^2 \middle| \begin{matrix}[f_1]\\ (\bar{\lambda}_1\bar{\mu}_1)\bar{S}_1\bar{\omega}_1\end{matrix} \otimes \begin{matrix}[f_2]\\ (\bar{\lambda}_2\bar{\mu}_2)\bar{S}_2\bar{\omega}_2\end{matrix};(\lambda\mu)S\omega \right\rangle, \tag{8}$$

where
$\delta_{1\bar{1}}\delta_{2\bar{2}} \equiv \delta((\lambda_1\mu_1),(\bar{\lambda}_1\bar{\mu}_1))\delta((\lambda_2\mu_2),(\bar{\lambda}_2\bar{\mu}_2))\delta(S_1\bar{S}_1)\delta(S_2\bar{S}_2)\delta(\omega_1\bar{\omega}_1)\delta(\omega_2\bar{\omega}_2)$.

The matrix element in Eq. (8) is evaluated by decoupling quark n_1 from the first group and quark n_2 from the second group of quarks by using the Wigner coefficients from Ref. 9, and recoupling using 9-j and SU(3) 9 $-\lambda\mu$ coefficients. The last term of Eq. (8) becomes

$$\sum_{\substack{(\lambda'_1\mu'_1)S'_1\omega'_1\\ [f'_1]}} \left\langle \begin{matrix}[f'_1]\\ (\lambda'_1\mu'_1)S'_1\omega'_1\end{matrix} \quad \begin{matrix}[1]\\ (10)\tfrac{1}{2}\end{matrix} \middle| \begin{matrix}[f_1]\\ (\lambda_1\mu_1)S_1\omega_1\end{matrix} \right\rangle \left\langle \begin{matrix}[f'_1]\\ (\lambda'_1\mu'_1)S'_1\omega'_1\end{matrix} \quad \begin{matrix}[1]\\ (10)\tfrac{1}{2}\end{matrix} \middle| \begin{matrix}[f_1]\\ (\bar{\lambda}_1\bar{\mu}_1)\bar{S}_1\bar{\omega}_1\end{matrix} \right\rangle \frac{n'_{f_1}}{n_{f_1}}$$
$$\times \sum_{\substack{(\lambda'_2\mu'_2)S'_2\omega'_2\\ [f'_2]}} \left\langle \begin{matrix}[f'_2]\\ (\lambda'_2\mu'_2)S'_2\omega'_2\end{matrix} \quad \begin{matrix}[1]\\ (10)\tfrac{1}{2}\end{matrix} \middle| \begin{matrix}[f_2]\\ (\lambda_2\mu_2)S_2\omega_2\end{matrix} \right\rangle \left\langle \begin{matrix}[f'_2]\\ (\lambda'_2\mu'_2)S'_2\omega'_2\end{matrix} \quad \begin{matrix}[1]\\ (10)\tfrac{1}{2}\end{matrix} \middle| \begin{matrix}[f_2]\\ (\bar{\lambda}_2\bar{\mu}_2)\bar{S}_2\bar{\omega}_2\end{matrix} \right\rangle \frac{n'_{f_2}}{n_{f_2}}$$
$$\times \sum_{(\lambda'\mu')(\lambda_{23}\mu_{23})\rho'} U\begin{pmatrix}(\lambda'_1\mu'_1) & (10) & (\lambda_1\mu_1)\\ (\lambda'_2\mu'_2) & (10) & (\lambda_2\mu_2)\\ (\lambda'\mu') & (\lambda_{23}\mu_{23}) & (\lambda\mu)\end{pmatrix}_{\rho\rho'} U\begin{pmatrix}(\lambda'_1\mu'_1) & (10) & (\bar{\lambda}_1\bar{\mu}_1)\\ (\lambda'_2\mu'_2) & (10) & (\bar{\lambda}_2\bar{\mu}_2)\\ (\lambda'\mu') & (\lambda_{23}\mu_{23}) & (\lambda\mu)\end{pmatrix}_{\bar{\rho}\rho'}$$
$$\times \sum_{S'S_{23}} U\begin{pmatrix}S'_1 & \frac{1}{2} & S_1\\ S'_2 & \frac{1}{2} & S_2\\ S' & S_{23} & S\end{pmatrix} U\begin{pmatrix}S'_1 & \frac{1}{2} & \bar{S}_1\\ S'_2 & \frac{1}{2} & \bar{S}_2\\ S' & S_{23} & S\end{pmatrix} \left\langle (\lambda_{23}\mu_{23})S_{23} \middle| \sum_\mu [\alpha_\mu(n_1)+\alpha_\mu(n_2)]^2 - \sum_\mu \alpha_\mu(n_1)^2 - \sum_\mu \alpha_\mu(n_2)^2 \middle| (\lambda_{23}\mu_{23})S_{23} \right\rangle, \tag{9}$$

where n_f is the dimension of the representation $[f]$ of the symmetric group. The value of the two quark matrix element in Eq. (9) is $\mathscr{C}_6^{(2)}(q^2) - 2{\cdot}35/6$.

The sum over the 9 $-j$ and 9 $-\lambda\mu$ symbols may be simplified. Noting that the two-particle SU(6) Casimir operator assumes the values given in Table I, and noting which SU(3) ⊗ SU(2) representations are contained in each SU(6) representation, the expectation value of the two-quark Casimir operator may be rewritten as

$$\tfrac{34}{3} + 2(-)^{\mu_{23}+S_{23}+1} \tag{10}$$

The 34/3 may be combined with the $-35/3$ arising from the one quark Casimir operator appearing in Eq. (9). The sum over S', S_{23} and $(\lambda'\mu')$, (λ_2,μ_{23}) may be performed explicitly using the sum rules

$$\sum_{S'S_{23}} (-)^{S_{23}} U\begin{pmatrix}S'_1 & \frac{1}{2} & S_1\\ S'_2 & \frac{1}{2} & S_2\\ S' & S_{23} & S\end{pmatrix} U\begin{pmatrix}S'_1 & \frac{1}{2} & \bar{S}_1\\ S'_2 & \frac{1}{2} & \bar{S}_2\\ S' & S_{23} & S\end{pmatrix} = (-)^{2S'_2+S_2+\bar{S}_2} U\begin{pmatrix}S'_1 & \frac{1}{2} & S_1\\ \frac{1}{2} & S'_2 & S_2\\ \bar{S}_1 & \bar{S}_2 & S\end{pmatrix} \tag{11}$$

$$\sum_{(\lambda'\mu')(\lambda_2\mu_{23})\rho'} (-)^{\mu_{23}} U\begin{pmatrix}(\lambda'_1\mu'_1) & (10) & (\lambda_1\mu_1)\\ (\lambda'_2\mu'_2) & (10) & (\lambda_2\mu_2)\\ (\lambda'\mu') & (\lambda_{23}\mu_{23}) & (\lambda\mu)\end{pmatrix}_{\rho'\rho} U\begin{pmatrix}(\lambda'_1\mu'_1) & (10) & (\bar{\lambda}_1\bar{\mu}_1)\\ (\lambda'_2\mu'_2) & (10) & (\bar{\lambda}_2\bar{\mu}_2)\\ (\lambda'\mu') & (\lambda_{23}\mu_{23}) & (\lambda\mu)\end{pmatrix}_{\rho'\bar{\rho}}$$
$$= (-)^{\lambda_2+\mu_2+\bar{\lambda}_2+\bar{\mu}_2} U\begin{pmatrix}(\lambda'_1\mu'_1) & (10) & (\lambda_1\mu_1)\\ (10) & (\lambda'_2\mu'_2) & (\lambda_2\mu_2)\\ (\bar{\lambda}_1\bar{\mu}_1) & (\bar{\lambda}_2\bar{\mu}_2) & (\lambda\mu)\end{pmatrix}_{\rho\bar{\rho}}. \tag{12}$$

The SU(2) sum rule, Eq. (11), was originally demonstrated in Ref. 10.

The matrix element of Eq. (9) may finally be written as

$$\delta_{1\bar{1}}\delta_{2\bar{2}}\left\{\mathscr{C}_6^{(2)}([f_1])+\mathscr{C}_6^{(2)}([f_2])-\tfrac{1}{3}n_1n_2\right\}+2(-)^{S_2+\bar{S}_2+\lambda_2+\mu_2+\bar{\lambda}_2+\bar{\mu}_2}\sum\left\langle\begin{matrix}[f_1'] & [1] \\ (\lambda_1'\mu_1')S_1'\omega_1' & (10)\tfrac{1}{2}\end{matrix}\middle|\begin{matrix}[f_1] \\ (\lambda_1\mu_1)S_1\omega_1\end{matrix}\right\rangle$$
$$\times\left\langle\begin{matrix}[f_1'] & [1] \\ (\lambda_1'\mu_1')S_1'\omega_1' & (10)\tfrac{1}{2}\end{matrix}\middle|\middle\rangle\begin{matrix}[f_1] \\ (\bar{\lambda}_1\bar{\mu}_1)\bar{S}_1\bar{\omega}_1\end{matrix}\right.\frac{n_{f_1'}}{n_{f_1}}\sum\frac{n_{f_2'}}{n_{f_2}}\left\langle\begin{matrix}[f_2'] & [1] \\ (\lambda_2'\mu_2')S_2'\omega_2' & (10)\tfrac{1}{2}\end{matrix}\middle|\begin{matrix}[f_2] \\ (\lambda_2\mu_2)S_2\omega_2\end{matrix}\right\rangle$$
$$\left\langle\begin{matrix}[f_2'] & [1] \\ (\lambda_2'\mu_2')S_2'\omega_2' & (10)\tfrac{1}{2}\end{matrix}\middle|\begin{matrix}[f_2] \\ (\bar{\lambda}_2\bar{\mu}_2)\bar{S}_2\bar{\omega}_2\end{matrix}\right\rangle(-)^{2S_2'}U\begin{pmatrix}S_1' & \frac{1}{2} & S_1\\ \frac{1}{2} & S_2' & S_2\\ \bar{S}_1 & \bar{S}_2 & S\end{pmatrix}U\begin{pmatrix}(\lambda_1'\mu_1') & (10) & (\lambda_1\mu_1)\\ (10) & (\lambda_2'\mu_2') & (\lambda_2\mu_2)\\ (\bar{\lambda}_1\bar{\mu}_1) & (\bar{\lambda}_2\bar{\mu}_2) & (\lambda\mu)\end{pmatrix}_{\rho\bar{\rho}}. \tag{13}$$

TABLE II. The Wigner coefficients for coupling two SU(6) representations [3] to a final SU(6) representation $[f]$ with subgroup SU(3) ⊗ SU(2). The SU(3) labels $(\lambda\mu)$ are defined by $\lambda = g_1 - g_2$ and $\mu = g_2 - g_3$.

		[3]×[3]		
$[f](\lambda\mu)S$	$(\Sigma 0)\frac{3}{2}\times(30)\frac{3}{2}$	$(30)\frac{3}{2}\times(11)\frac{1}{2}$	$(11)\frac{1}{2}\times(30)\frac{3}{2}$	$(11)\frac{1}{2}\times(11)\frac{1}{2}$
[42] (11)0				1
[51] (11)7		$-\sqrt{2}/3$	$\frac{1}{3}\sqrt{2}$	$\frac{1}{3}\sqrt{5}$
[42] (11)1		$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	
[411] (11)1		$\sqrt{\frac{5}{18}}$	$-\sqrt{\frac{5}{18}}$	$\frac{2}{3}$
[42] (11)2		$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$	
[411] (11)2		$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	
[42] (30)0				1
[51] (30)1	$\frac{1}{\sqrt{18}}$	$\frac{1}{\sqrt{18}}$	$\sqrt{\frac{8}{9}}$	
[42] (30)1	$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$		
[411] (30)1	$\frac{2}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$	
[51] (30)2		$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$	
[42] (30)2		$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	
[6] (03)0	$\frac{1}{\sqrt{5}}$			$\sqrt{\frac{4}{5}}$
[42] (03)0	$\sqrt{\frac{4}{5}}$			$-\frac{1}{\sqrt{5}}$
[51] (03)1	$\sqrt{\frac{5}{9}}$			$\frac{2}{3}$
[411] (03)1	$\frac{2}{3}$			$-\frac{1}{3}\sqrt{5}$
[42] (03)2	1			
[33] (03)3	1			

TABLE II. (Continued).

[51] (22)0	$\frac{1}{3}\sqrt{5}$			$\frac{2}{3}$
[411] (22)0	$-\frac{2}{3}$			$\frac{1}{3}\sqrt{5}$
[6] (22)1	$\frac{1}{3}$	$\frac{1}{3}\sqrt{2}$	$\frac{1}{3}\sqrt{2}$	$\frac{2}{3}$
[51] (22)1		$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$	
[42] (22)1		$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{\sqrt{2}}$
[42] (22)1	$-\frac{2}{3}\sqrt{2}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}\sqrt{2}$
[51] (22)2	$\frac{2}{3}$	$\sqrt{\frac{5}{18}}$	$\sqrt{\frac{5}{18}}$	
[42] (22)2		$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$	
[411] (22)2	$\frac{1}{3}\sqrt{5}$	$-\frac{1}{3}\sqrt{2}$	$-\frac{1}{3}\sqrt{2}$	
[42] (22)3	1			
[411] (00)0				1
[411] (00)1				1

The SU(6) ⊃ SU(3) ⊗ SU(2) Wigner coefficients may be taken from Ref. 9. The unitary $9-j$ and $9-\lambda\mu$ coefficients may be calculated using available programs.[11] There is no outer multiplicity in the $(\lambda\mu)$ coupling save for that which already occurs in the definition of the product states.

Upon diagonalization of the resulting arrays, the eigenvalues will correspond to allowed values of the SU(6) Casimir operator, and the components of the eigenvectors are the desired SU(6) ⊃ SU(3) ⊗ SU(2) Wigner coefficients.

The above procedure defines the necessary steps if both wavefunctions in the basis state, Eq. (3), are comprised of quarks. However, if one cofiguration consists of antiquarks, e.g., $\bar{q}^{n_2}$, the procedure must be modified very slightly. If the group conjugate to SU(6) is SU(m), e.g., $SU(3)_f$, then a configuration of $6m$ quarks with SU(6), SU(3), and SU(2) representation [$mmmmmm$] (00)0 will have the same group properties as the vacuum and a state $|q^{6m-n}\,[m-f_6\cdots m-f_1](\lambda\mu)S\rangle$ will have the same group properties as n antiquarks. Hence, it suffices to substitute everywhere in Eq. (13) $6m-n_2$ for n_2. For the current problem this would require a knowledge of the SU(6) Wigner coefficients relating 17 quarks to 16 quarks. Since columns of six transform as a scalar under SU(6) transformations, and the corresponding Wigner coefficients are identical, it is adequate to know the coefficients relating 11 quarks to 10 quarks.

The techniques used above will work for any other groups although the use of the sum rules which simplify Eq. (14) may not be general. They will apply, however, to SU(4) ⊃ SU(2) ⊗ SU(2), and probably to SU(mn) ⊃ SU(m) ⊗ SU(n). In the event that there occur two or more representations of SU(mn) which have the same value for $\mathscr{C}_m^{(2)}$, the degeneracy may be lifted by diagonalizing $\mathscr{C}_m^{(3)}$, the cubic invariant, within the subspace of degenerate representations.

As an application of this procedure the SU(6) Wigner coefficients relevant for $q^3\times q^3$ have been calculated in those cases for which the SU(3) representation of the six quark state is (00) (e.g., a color singlet), (11) or octet, and for (30) or (03) decuplets. The coefficients are given in Table II.

ACKNOWLEDGMENT

I should like to thank D.J. Millener for a version of his code for calculating SU(3) $9-\lambda\mu$ symbols.

[1] J.J.J. Kokkedee, *The Quark Model* (Benjamin, New York, 1969).
[2] H.J. Melosh, Phys. Rev. D **9**, 1095 (1974).
[3] F.J. Gilman, M. Kugler, and S. Meshkov, Phys. Rev. D **9**, 715 (1974).
[4] F.J. Gilman and I. Karliner, Phys. Rev. D **10**, 2194 (1974).
[5] Y. Akiyama and J.D. Draayer, Math. Phys. **14**, 1904 (1973).
[6] C.L. Cook and G. Murtzaz, Nuovo Cimento **39**, 7095 (1965); M. Machacek and Y. Tomozawa, J. Math. Phys. **17**, 458 (1976).
[7] R.L. Jaffee, Phys. Rev. **D14**, 267 (1977).
[8] A. Chodos, R.L. Jaffee, K. Johnson, and C. Thorn, Phys. Rev. D **10**, 2599 (1974).
[9] S.I. So and D. Strottman, J. Math. Phys. **20**, 153 (1976).
[10] A. Arima, H. Horie, and Y. Tanabe, Prog. Theor. Phys. (Kyoto) **11**, 143 (1954).
[11] D.J. Millener, J. Math. Phys. **19**, 1513 (1978).

Singular spin harmonics

G. J. Weir

Applied Mathematics Division, D.S.I.R., Wellington, New Zealand
(Received 6 September 1978)

The definition of the spin-weighted spherical harmonics, ${}_sY_{lm}$, is extended so that all integral spin weights, s, are allowed. We then discuss a set of spin-weighted functions, ${}_sZ_{lm}$, which are analogous to the spherical harmonics of the second kind.

1. INTRODUCTION

A fruitful area for research has been provided recently by the spin-weighted spherical harmonics, which have been discussed in relation to representation theory,[1] conserved quantities in relativity theory,[2] Maxwell's equations,[3] the gravitational field equations,[4] and in manifold theory.[5] In all applications to date the fields being described by spin-weighted spherical harmonics (hereafter referred to as spin harmonics) have been regular[6] on all points of the unit sphere.

This paper relaxes the regularity condition, thereby allowing nonregular points on the unit sphere, which are necessary if the formalism of spin harmonics is to describe, for example, the half-space problems of geophysics, where singularities occur in the magnetic field at the positions of current electrodes. Only integral spin fields are discussed in this paper.

2. SPIN HARMONICS

In this section we review some results which are needed later.

For a field, ζ, of spin weight s, the differential operators, $\eth$ and $\bar{\eth}$, are defined by

$$\eth\zeta = -(\partial_\theta + i\csc\theta\partial_\phi - s\cot\theta)\zeta, \tag{2.1}$$

$$\bar{\eth}\zeta = -(\partial_\theta - i\csc\theta\partial_\phi + s\cot\theta)\zeta, \tag{2.2}$$

where (r,θ,ϕ) is a spherical polar coordinate system. The spin harmonics, ${}_sY_{lm}$, are defined as

$${}_sY_{lm} = e^{im\phi}\left(\sin\frac{\theta}{2}\right)^{2l}\sum_n a_{slmn}\left(\cot\frac{\theta}{2}\right)^{2n+s-m}, \tag{2.3}$$

$$a_{slmn} = (-1)^{l-n-s}\left(\frac{(l+m)!(l-m)!(2l+1)}{(l+s)!(l-s)!4\pi}\right)^{1/2} \times C_n^{l-s}C_{n+s-m}^{l+s}, \tag{2.4}$$

and satisfy[7]

$${}_0Y_{lm} = Y_{lm}, \quad {}_{-s}Y_{lm} = (-1)^{s+m}\,{}_s\bar{Y}_{l-m}, \tag{2.5}$$

$$\eth\,{}_sY_{lm} = \sqrt{(l-s)(l+s+1)}\,{}_{s+1}Y_{lm}, \tag{2.6}$$

$$\bar{\eth}\,{}_sY_{lm} = -\sqrt{(l+s)(l-s+1)}\,{}_{s-1}Y_{lm}, \tag{2.7}$$

$$\bar{\eth}\eth\,{}_sY_{lm} = (s-1)(l+s+1)\,{}_sY_{lm}, \tag{2.8}$$

where $|s| \leqslant 1$ and $l \geqslant |m|$.

Goldberg *et al.*[8] have shown that the ${}_sY_{lm}$ form a complete orthonormal set of functions for regular integral spin fields on the unit sphere. However, the ${}_sY_{lm}$ are not defined for $|s| > l$, and so from Eq. (2.7), they cannot describe field, ξ, which satisfies $\bar{\eth}\xi = {}_0Y_{00}$. This is shown in the Appendix to be the equation for the magnetic field from a steady state electric monopole in a uniform half-space, which suggests extending the definition of the spin harmonics to allow for all integral values of s.

3. SINGULAR SPIN HARMONICS

The singular spin harmonics ${}_sY_{lm}$ are defined as

$${}_sY_{lm} = e^{im\phi}\left(\tan\frac{\theta}{2}\right)^{s-m}\sum_{n=0}^{l-m} b_{slmn}\left(\sin\frac{\theta}{2}\right)^{2m+2n}, \tag{3.1}$$

$$b_{slmn} = (-1)^{s+n-l}\left(\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right)^{1/2} \times \frac{(m+l+n)!}{(m+s+n)!n!(l-m-n)!}, \tag{3.2}$$

and satisfy

$${}_{-s}Y_{lm} = (-1)^{m+s}\,{}_s\bar{Y}_{l-m}, \tag{3.3}$$

$$\eth\,{}_sY_{lm} = (s-l)(l+s+1)\,{}_{s+1}Y_{lm}, \tag{3.4}$$

$$\bar{\eth}\,{}_sY_{lm} = {}_{s-1}Y_{lm}, \tag{3.5}$$

$$\eth\,{}_{-s}Y_{lm} = {}_{-s+1}Y_{lm}, \tag{3.6}$$

$$\bar{\eth}\,{}_{-s}Y_{lm} = (s-l)(l+s+1)\,{}_{-s-1}Y_{lm}, \tag{3.7}$$

with ${}_sY_{lm}$ being regular at the north pole ($\theta = 0$), $s,l,|m|$ being nonnegative integers with $s > l \geqslant |m|$, and ${}_lY_{lm}$, ${}_{-l}Y_{lm}$ being the spin harmonics defined in Eqs. (2.3) and (2.4).

Since the ${}_sY_{lm}$ are not square integrable over the unit sphere when $|s| > l$, the definitions just given could be modified by constant normalization factors, which would alter the numerical coefficients in Eqs. (3.3)–(3.7). Singular spin harmonics can also be defined via the hypergeometric functions[9] or through Eqs. (3.5) and (3.6), together with a regularity condition either at the north pole, or at the south pole ($\theta = \pi$).

4. THE SPIN-WEIGHTED FUNCTIONS

Newman and Penrose have shown that $\eth^s{}_0Y_{lm}$ and $\bar{\eth}^s{}_0Y_{lm}$ have spin weights s and $-s$, respectively. The spherical harmonics of the second kind, ${}_0Z_{lm}$, which are defined[11] as

$${}_0Z_{lm} = \left(\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right)^{1/2} Q_l^m(\cos\theta)e^{im\phi}, \tag{4.1}$$

$$Q_l^m(x) = (-1)^m(1-x^2)^{m/2}\frac{d^m}{dx^m}Q_l(x), \tag{4.2}$$

 0022-2488/79/081648-02$01.00

$$Q_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}\left((x^2-1)^l \operatorname{Ln}\frac{1+x}{1-x}\right) - \frac{1}{2}P_l(x)\operatorname{Ln}\frac{1+x}{1-x}, \tag{4.3}$$

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}[(x^2-1)^l], \tag{4.4}$$

transform as scalars, and so the quantities[12]

$$_{|s|}Z_{lm} = \eth^{|s|}{}_0Z_{lm}, \tag{4.5}$$

$$_{-|s|}Z_{lm} = \bar{\eth}^{|s|}{}_0Z_{lm}, \tag{4.6}$$

have spin weights s and $-s$ respectively. As with the singular spin harmonics, the definitions given above could be altered by normalization factors since the ${}_sZ_{lm}$ are not square integrable over the unit sphere.

From the definition of ${}_0Z_{lm}$,

$$\bar{\eth}\eth\,{}_0Z_{lm} = -l(l+1){}_0Z_{lm}$$

or

$$\bar{\eth}_1Z_{lm} = -l(l+1){}_0Z_{lm}.$$

Using the commutator relation

$$[\bar{\eth},\eth] = 2s, \tag{4.7}$$

and Eqs. (4.5) and (4.6), it can be shown by induction that, for positive s,

$$\bar{\eth}_s Z_{lm} = (l+s)(s-l-1)_{s-1}Z_{lm}, \tag{4.8}$$

$$\eth_{-s}Z_{lm} = (l+s)(s-l-1)_{-s+1}Z_{lm}, \tag{4.9}$$

$$\bar{\eth}\eth_s Z_{lm} = (s-l)(l+s+1)_s Z_{lm}. \tag{4.10}$$

We shall now show that the ${}_sZ_{lm}$ are nonzero for all integral s, and that for fixed s, l, m, the ${}_sY_{lm}$ and ${}_sZ_{lm}$ are linearly independent. For this it is sufficient to consider only positive s.

Let $r(l,m)$ be the first integer, for fixed l and m, for which ${}_{r+1}Z_{lm}$ is zero. The linearity of $\eth$ in Eq. (2.1) then shows that ${}_rZ_{lm}$ is proportional to ${}_lY_{lm}$, so that $r = l$. By hypothesis, ${}_sZ_{lm}$ is nonzero for all $s < r+1$, and so applying $\bar{\eth}^l$ to both ${}_lY_{lm}$ and ${}_lZ_{lm}$, and using Eq. (4.8), shows that ${}_0Y_{lm}$ and ${}_0Z_{lm}$ are proportional. This contradiction proves that ${}_sZ_{lm}$ are nonzero for all s. Linear independence follows by operating repeatedly with $\bar{\eth}$, together with the facts that firstly ${}_0Y_{lm}$ and ${}_0Z_{lm}$ are linearly independent, and secondly that ${}_lY_{lm}$ and ${}_lZ_{lm}$ cannot be linearly dependent since $\eth\,{}_lY_{lm}$ is zero, whereas $\eth\,{}_lZ_{lm}$ is nonzero.

A geometric picture of the action of $\eth$ and $\bar{\eth}$ on the ${}_sY_{lm}$ and ${}_sZ_{lm}$ can be obtained by associating with each pair $({}_sY_{lm}, {}_sZ_{lm})$ the point (m,l,s) in a right-handed Cartesian coordinate system, with positive s along the positive vertical direction. We call the set of points with $|s| \leqslant l$ and $0 \leqslant |m| \leqslant l$ the inner region I and those with $|s| > l \geqslant |m| \geqslant 0$ the outer region. The set of points in the outer region for which $s > 0$ and $s < 0$ are denoted by θ^+ and θ^-, respectively. Then $\eth$ and $\bar{\eth}$ allow us to move vertically up and down in I, in θ^+, or in θ^-. However, we can only cross from the inner to the outer region along the ${}_sZ_{lm}$, and only from the outer to the inner regions along ${}_sY_{lm}$.

Finally, the notation above was chosen so that ${}_sY_{lm} \sim {}_sP_l^m e^{im\phi}$ and ${}_sZ_{lm} \sim {}_sQ_l^m e^{im\phi}$. Unfortunately, this notation clashes with that of Teukolsky, who denoted by spheroidal harmonics (${}_{as}Y_{lm}$?) by ${}_sZ_{lm}$.

APPENDIX

A steady state electric monopole denotes a steady current I, flowing up from the south pole of a spherical polar coordinate system, and into a uniform half-space, $0 \leqslant \theta \leqslant \pi/2$, of constant conductivity σ. The magnetic fields have only a ϕ component, and so the quantities $\pm iB_\phi$ have spin weights ± 1. For $\pi \geqslant \theta \geqslant \pi/2$, $B_\phi = \mu I \csc\theta/2\pi r$, while for $\pi/2 \geqslant \theta \geqslant 0$, $B_\phi = \mu I \tan(\frac{1}{2}\theta)/2\pi r$. It is straightforward to show that in the conducting medium, Maxwell's equations

$$\eth B_0 = -\partial_r(rB_+) - ir(\mu\sigma E_+ + \mu\epsilon\partial_t E_+),$$

$$\eth E_0 = ir\partial_t B_+ - \partial_r(rE_+),$$

$$\bar{\eth}(rB_+) = \partial_r(r^2B_0) - ir^2(\mu\sigma E_0 + \mu\epsilon\partial_t E_0),$$

$$\bar{\eth}(rE_+) = \partial_r(r^2E_0) + ir^2\partial_t B_0,$$

$$B_0 = B_r, \quad E_0 = E_r,$$

$$B_+ = B_\theta + iB_\phi, \quad E_+ = E_\theta + iE_\phi,$$

imply

$$\bar{\eth}\xi = {}_0Y_{00}, \quad \text{where } B_\phi = -\mu I\xi/\sqrt{\pi}\, r.$$

[1] I.M. Gel'fand, R.A. Milnos, and Z.Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Macmillan, New York, 1963).
[2] E.T. Newman and R. Penrose, J. Math. Phys. **7**, 862 (1966).
[3] A. Janis and E.T. Newman, J. Math. Phys. **6**, 902 (1965).
[4] L. Derry, R. Isaacson, and J. Winicour, Phys. Rev. **5**, 1647 (1969).
[5] W.D. Curtis and D.E. Lerner, J. Math. Phys. **19**, 847 (1978).
[6] Ref. 3, p. 867.
[7] R.A. Breuer, M.P. Ryan, and S. Waller, Proc. Roy. Soc. A **358**, 71 (1977).
[8] J.N. Goldberg, A.J. MacFarlane, E.T. Newman, F. Rohrlich, and E.G.G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).
[9] Ref. 7, Eq. (3.7).
[10] S. Teukolsky, Ap. J. **185**, 635 (1973). Our ${}_sZ_{lm}$ are not the spheroidal harmonics of Teukolsky.
[11] W. Magnus and F. Oberhettinger, *Functions of Mathematical Physics* (Chelsea, New York, 1949), pp. 58 and 60.
[12] The magnetic field in the air from a steady state electric monopole is associated with the spinfield, ${}_1Z_{00}$, since $\sqrt{4\pi}\,{}_1Z_{00} = \csc\theta$.

An application of Lax's estimates to the determination of critical crystal thickness in nonlinear laser optics

Piero Bassanini

Istituto Matematico, Università di Perugia, Italy
(Received 11 December 1978)

Lax's estimates concerning the development of singularities of solutions of the Cauchy problem for 2×2 quasilinear hyperbolic systems are applied to the Maxwell system of nonlinear optics, which describes the propagation of a plane polarized laser wave through a uniaxial piezoelectric crystal, in a direction orthogonal to the optic axis, and neglecting absorption and dispersion. The "critical thickness" of the crystal, i.e., the distance after which (shock) singularities necessarily occur, is evaluated within an error of order less than 10^{-3}, and is found to be proportional to the ratio of laser wavelength to peak amplitude times the nonlinear optical coefficient. The critical thickness appears to be ordinarily of the order of a few cm.

INTRODUCTION

We consider the following Cauchy problem for the 2×2 quasilinear hyperbolic Maxwell system:

$$\frac{\partial E}{\partial x}+\mu_0\frac{\partial H}{\partial t}=0,$$
$$\frac{\partial H}{\partial x}+\epsilon_2\frac{\partial E}{\partial t}+2\eta E\frac{\partial E}{\partial t}=0, \tag{1}$$

for $t\in\mathbb{R}$, $x>0$, and

$$E(0,t)=2(1+K)^{-1}E_L\cos(\omega t),$$
$$H(0,t)=2K(1+K)^{-1}(\epsilon_0/\mu_0)^{1/2}E_L\cos(\omega t), \tag{2}$$

for $t\in\mathbb{R}$, $x=0$, where $K=(\epsilon_2/\epsilon_0)^{1/2}$. This problem corresponds to a mathematically idealized version of the following physical situation[1-7]: A plane polarized (monochromatic harmonic) laser wave, with wavelength $\lambda=2\pi c_0/\omega$ $[c_0=(\epsilon_0\mu_0)^{-1/2}]$ and peak amplitude E_L, propagating in vacuo along the x axis, is normally incident on a crystal slab $D_a=\{0<x<a\}$ $(a>0)$ with refractive index K and nonlinear optical coefficient proportional to η [see (5) and (11)]. Then, letting $a\to+\infty$ yields the half-space problem (1), (2).

It is well known that if the crystal is piezoelectric, (second) harmonic generation may occur for the reflected $(x<0)$ and transmitted $(x>a)$ wave: This phenomenon was first revealed for quartz by the experiments of Franken *et al.*[8] and has since been the object of a large amount of experimental and theoretical investigations (see, e.g., Refs. 9 and 10) using crystals of thickness of the order of 1 cm, or less.

In some recent papers (Refs. 1–6; see also Refs. 7, 11, and 12) a rigorous mathematical theory of (second) harmonic generation of laser radiation by a (thin) slab D_a of a piezoelectric crystal has been proposed. This theory is based on the study of a boundary value problem (at $x=0,a$) for the system (1), and is also physically rigorous insofar as absorption and dispersion can be neglected, the crystal belongs to suitable uniaxial classes, and the geometry is conveniently chosen: in particular if the crystal is of class 32 — D3 with optic axis parallel to the magnetic field H (and orthogonal to the x axis and to the electric field E).[11,12] Then, the presence of a second harmonic wave has been rigorously proved, in the framework of this theory, by means of a convergent numerical method of successive approximations,[5-6] based on the existence and uniqueness theorem given by Cesari,[3-4] provided a is small enough, $a<\bar{a}$; an estimate by defect of $\bar{a}$ is included in the proof ($\bar{a}\simeq0.2$ cm. for quartz, see Ref. 7).

When $a\to+\infty$, the linear reflected wave for $x>0$ disappears, and the boundary conditions reduce to Cauchy initial conditions for $x=0$. (See Sec. 3.) Then, according to a general result by P.D. Lax,[13] the smooth solution of (1) and (2) does not exist for values of x exceeding a "critical value" a_c (Theorem 1), and estimates by excess and by defect of a_c can be obtained by applying to (1) and (2) the general estimates of Ref. 13, as explicitly developed in Ref. 14. The estimates by excess and by defect thus derived are found to coincide at the order of approximation $O(E_L\chi_2)$ [see (14)], and yield therefore an *exact* value of the critical thickness a_c at the same order of approximation [see (13)].

2. ESTIMATE OF THE CRITICAL THICKNESS

We shall follow the notations of Refs. 7 and 14.

Let us define, as usual, $c_2=(\epsilon_2\mu_0)^{-1/2}$, $K=c_0/c_2$; next, we define the new independent variable $y=c_2t$ and new unknown functions $u=u(x,y)$, $v=v(x,y)$ as follows:

$$K(\epsilon_0/\mu_0)^{1/2}E(x,t)=u(x,y),\qquad H(x,t)=v(x,y).$$

Then, (1) and (2) become

$$u_x+v_y=0,\qquad v_x+u_y+2\epsilon uu_y=0\qquad (y\in\mathbb{R},\ x>0), \tag{3}$$

$$u(0,y)=\Omega\cos(2\pi K\lambda^{-1}y),\qquad v(0,y)=u(0,y)\qquad (y\in\mathbb{R},\ x=0), \tag{4}$$

where

[a]Research supported by the GNFM of CNR.

 0022-2488/79/081650-03$01.00

$$\epsilon = (\epsilon_0 K^2)^{-1}\mu_0 c_2 \eta, \quad \Omega = 2\alpha_0 E_L K (K+1)^{-1},$$

$$\alpha_0 = \left(\frac{\epsilon_0}{\mu_0}\right)^{1/2}. \tag{5}$$

System (3) is of the form $w_x + A(w)w_y = 0$, $w = (u,v)$, and matrix

$$A = [0,\ 1;\ 1 + 2\epsilon u,\ 0].$$

Eigenvalues and left eigenvectors of A are given by

$$\rho_1 = (1 + 2\epsilon u)^{1/2} = -\rho_2, \quad b_1 = (\rho_1, 1), \quad b_2 = (\rho_2, 1). \tag{6}$$

Thus, (3) is *strictly hyperbolic* in the sense of Lax[15] provided $2|\epsilon u| < 1$ [see (14)], and can be reduced to the "second canonic form"[3] by taking scalar products by b_1 and b_2, respectively,

$$(v_x + \rho_i v_y) + \rho_i(u_x + \rho_i u_y) = 0 \quad (i = 1,2).$$

The Riemann invariants z_1, z_2 satisfy the equations (see Ref. 16, p. 430):

$$\frac{\partial z_i}{\partial u} = \rho_i, \quad \frac{\partial z_i}{\partial v} = 1, \quad (i = 1,2),$$

whence

$$z_1 = v + (3\epsilon)^{-1}(1 + 2\epsilon u)^{3/2} - (3\epsilon)^{-1},$$
$$z_2 = v - (3\epsilon)^{-1}(1 + 2\epsilon u)^{3/2} + (3\epsilon)^{-1},$$

System (3) can now be written in the "first canonic form"[3,14] in terms of the Riemann invariants $(z_1, z_2) = z$:

$$\frac{\partial z_1}{\partial x} + \rho_1(z)\frac{\partial z_1}{\partial y} = 0,$$
$$\frac{\partial z_2}{\partial x} + \rho_2(z)\frac{\partial z_2}{\partial y} = 0, \tag{7}$$

with Cauchy conditions

$$z_1(0,y) \equiv \varphi_1(y) = \Omega\cos(2\pi K\lambda^{-1}y) + (3\epsilon)^{-1}[1 + 2\epsilon\Omega\cos(2\pi K\lambda^{-1}y)]^{3/2} - (3\epsilon)^{-1}, \tag{7'}$$

$$z_2(0,y) \equiv \varphi_2(y) = \Omega\cos(2\pi K\lambda^{-1}y) - (3\epsilon)^{-1}[1 + 2\epsilon\Omega\cos(2\pi K\lambda^{-1}y)]^{3/2} + (3\epsilon)^{-1},$$

Since in general $|\epsilon\Omega| \ll 1$ [see Refs. 7, 9, and 10 and Eqs. (11) and (14)]: $|\epsilon\Omega| < 10^{-6}$, we shall henceforth neglect terms of the order of $o(\epsilon\Omega)$ [i.e., terms of the order $o(E_L\chi_2)$, see (14)] and take into consideration the approximate expressions of (7) and (7') up to and including terms of the order $O(E_L\chi_2)$. We then find:

$$z_1 \cong v + u + 2^{-1}(\epsilon u)u, \quad z_2 \cong v - u - 2^{-1}(\epsilon u)u, \tag{8}$$

$$\rho_1 \cong 1 + \epsilon u \cong 1 + 2^{-1}\epsilon(z_1 - z_2) \cong -\rho_2, \tag{9}$$

while for the Cauchy data we need only retain terms of order $O(1)$:

$$z_1(0,y) \cong v(0,y) + u(0,y) = 2\Omega\cos(2\pi K\lambda^{-1}y) \cong \varphi_1(y),$$
$$z_2(0,y) \cong v(0,y) - u(0,y) = 0 \cong \varphi_2(y). \tag{10}$$

Hence, $z_1 \in [-2\Omega, 2\Omega]$ (see Ref. 14).

Following the notation of Refs. 7 and 14 we have here:

$$\Lambda = 4\pi K\lambda^{-1}\Omega \equiv 8\pi K^2(K+1)^{-1}\lambda^{-1}(\epsilon_0/\mu_0)^{1/2}E_L,$$
$$L = |\epsilon| = 4\pi K^{-3}(c_0\epsilon_0)^{-1}|\chi_2|, \tag{11}$$

where $4\pi\chi_2$ is the relevant nonlinear optical coefficient (divided by ϵ_0) in the rationalized MKSA (Giorgi) system of units (χ_2 is expressed in m/Volt). We shall henceforth assume, without loss of generality, $\epsilon > 0$, $\chi_2 > 0$ [$\chi_2 > 0$ for quartz, see Ref. 10; the case of negative χ_2 can be recovered by a π-phase shift of the incident laser wave (2)]. From (9) we find

$$\frac{\partial\rho_1}{\partial z_1} \cong \frac{\partial\rho_2}{\partial z_2} \cong \frac{\epsilon}{2} \cong -\frac{\partial\rho_1}{\partial z_2} \cong -\frac{\partial\rho_2}{\partial z_1},$$
$$\rho_1 - \rho_2 = 2\rho_1 \cong 2(1 + \epsilon u) \cong 2; \tag{12}$$

thus, system (7) is *genuinely nonlinear* in the sense of Lax,[15] and all assumptions of Ref. 14 are satisfied.

Lax's estimates for strictly hyperbolic and genuinely nonlinear 2×2 systems of partial differential equations are based on evaluations of the following quantities[14]:

$$W_1 = \int_{z_2}^{\varphi_2(y)} d\alpha \left(\frac{\partial\rho_1(\varphi_1(y),\alpha)}{\partial\alpha}\right) \times [\rho_1(\varphi_1(y),\alpha) - \rho_2(\varphi_1(y),\alpha)]^{-1},$$

$$W_2 = \int_{\varphi_1(y)}^{z_1} d\alpha \left(\frac{\partial\rho_2(\alpha,\varphi_2(y))}{\partial\alpha}\right) \times [\rho_1(\alpha,\varphi_2(y)) - \rho_2(\alpha,\varphi_2(y))]^{-1},$$

$$A_1 = \sup_y\left\{-\left(\frac{d\varphi_1}{dy}\right)\min_{z_2}\left[\left(\frac{\partial\rho_1(\varphi_1(y),z_2)}{\partial z_1}\right)\exp(W_1)\right]\right\},$$

$$A_2 = \sup_y\left\{-\left(\frac{d\varphi_2}{dy}\right)\min_{z_1}\left[\left(\frac{\partial\rho_2(z_1,\varphi_2(y))}{\partial z_2}\right)\exp(W_2)\right]\right\},$$

$$B_1 = \sup_y\left\{\left|\frac{d\varphi_1}{dy}\right|\max_{z_2}\left[\left(\frac{\partial\rho_1(\varphi_1(y),z_2)}{\partial z_1}\right)\exp(W_1)\right]\right\},$$

$$B_2 = \sup_y\left\{\left|\frac{d\varphi_2}{dy}\right|\max_{z_1}\left[\left(\frac{\partial\rho_2(z_1,\varphi_2(y))}{\partial z_2}\right)\exp(W_2)\right]\right\},$$

where the sup with respect to y is taken on $\mathbb{R}$ and the max and min with respect to z_i are taken on $[-2\Omega, 2\Omega]$.[14]

Define $A_c = \max_i(A_i)$, $B_c = \max_i(B_i)$ $(i = 1,2)$. Then, Lax's theorem[13] can be stated as follows[14]:

Theorem 1: *The smooth (differentiable) solution of* (7), (7') *exists for all* $x < B_c^{-1}$, *and does not exist for* $x \geqslant A_c^{-1}$. (If $A_c \leqslant 0$, estimate A_c^{-1} drops out). See also Ref. 18.

Since $\varphi_2(y) \cong 0$, the solution of (7), (7') at the order of approximation stated above is a *simple wave*[16] ($z_2 \equiv 0$ for all $x \geqslant 0$). Then, according to Ref. 14):

$$a_c \geqslant (\Lambda L)^{-1}.$$

By force of (11) we find here:

$$a_c \geqslant (32\pi^2)^{-1}K(K+1)\lambda(E_L\chi_2)^{-1} \equiv (\Lambda L)^{-1}.$$

Moreover, since the solution is (approximately) a simple wave, we know (Ref. 14, Corollary 2) that Lax's estimate by excess A_c^{-1} yields the exact value of a_c: Hence

$$(\Lambda L)^{-1} \leqslant a_c = A_c^{-1}.$$

We shall now show that $A_c^{-1} = 2(\Lambda L)^{-1}$ [see (13)]. Indeed, by force of (8)–(12) we find here:

$$W_1 \cong -(\epsilon/2)\int_{z_2}^{0} d\alpha\,(1/2) = (\epsilon/4)z_2 = 0,$$

$$W_2 \cong -(\epsilon/2)\int_{\varphi_1(y)}^{z_1} d\alpha(1/2) = -(\epsilon/4)[z_1 - \varphi_1(y)]$$

$$A_1 \cong \sup_y\left[-\left(\frac{d\varphi_1}{dy}\right)\min_{z_2}\left(\frac{\epsilon}{2}\exp(W_1)\right)\right]$$

$$\cong \frac{\epsilon}{2}\sup_y\left[-\left(\frac{d\varphi_1}{dy}\right)\right] \cong \frac{\epsilon}{2}\Lambda,$$

see (11) and (14);

$$A_2 \cong \sup_y\left[-\left(\frac{d\varphi_2}{dy}\right)\min_{z_1}\left(\frac{\epsilon}{2}\exp(W_2)\right)\right] \cong 0;$$

$$B_1 \cong \sup_y\left[\left|\frac{d\varphi_1}{dy}\right|\max_{z_2}\left(\frac{\epsilon}{2}\exp(W_1)\right)\right]$$

$$\cong \frac{\epsilon}{2}\sup_y\left(\left|\frac{d\varphi_1}{dy}\right|\right) \cong A_1$$

$$B_2 \cong \sup_y\left[\left|\frac{d\varphi_2}{dy}\right|\max_{z_1}\left(\frac{\epsilon}{2}\exp(W_2)\right)\right] \cong 0.$$

Thus, we have $A_c \cong B_c \cong A_1 \cong \Lambda L/2 = (\epsilon/2)\Lambda$, the estimates by excess and by defect of a_c coincide (see also Theorem 6 of Ref. 14), and $a_c = 2(\Lambda L)^{-1}$.

Summarizing, we find here, at the order of approximation $O(E_L\chi_2)$, the *exact* estimate for the critical thickness of the crystals:

$$a_c = \frac{K(K+1)}{16\pi^2}\frac{\lambda}{E_L\chi_2}. \tag{13}$$

The nondimensional number $E_L\chi_2$ is usually very small in experiments[7,9,10]: If we assume

$$E_L\chi_2 < 10^{-6}, \tag{14}$$

we find that the critical thickness is of the order of *a few centimeters.*

For instance, in the case of quartz, according to Ref. 10, we have the values: $K \cong 1.54$, $\lambda \cong 7000$ Å, $\chi_2 \cong 2.8\times 10^{-14}$ m/Volt, $E_L\chi_2 \cong 2.8\times 10^{-7}$ whence $a_c \cong 6$ cm.

3. THE SECOND-HARMONIC REFLECTED WAVE

The Cauchy conditions (2) at the wall $x = 0$ are admittedly approximate ones, valid at the order of approximation $O(1)$ (which is sufficient for the purposes of Sec. 2) and do not allow for the presence of a (second-) harmonic reflected wave for $x < 0$. The exact Cauchy data for $x = 0$ can be found under the assumption (see the introduction) that the solution inside the half-space $x \geqslant 0$ be (rigorously) given by the nonlinear progressive simple wave $z_1(x,y)$, i.e., $z_2(x,y) \equiv 0$ for $x \geqslant 0$. For this it is necessary and sufficient that $\varphi_2(y) \equiv 0$, that is

$$v(0,y) - (3\epsilon)^{-1}[1 + 2\epsilon u(0,y)]^{3/2} + (3\epsilon)^{-1} = 0,$$

[see the formulas preceding (7)], whence

$$H(0,t) - (3\epsilon)^{-1}\,[1 + 2\epsilon K(\epsilon_0/\mu_0)^{1/2}E(0,t)]^{3/2} + (3\epsilon)^{-1} = 0. \tag{15}$$

On the other hand, since for $x < 0$ the e.m. field satisfies the linear Maxwell system [i.e., system (1) with $\eta = 0$ and ϵ_2 replaced by ϵ_0] whose general integral is well known in terms of two arbitrary opposite travelling (linear) simple waves, by the same token as in[1-7] we find:

$$(\epsilon_0/\mu_0)^{1/2}E(0,t) + H(0,t) = 2(\epsilon_0/\mu_0)^{1/2}E_L\cos(\omega t). \tag{16}$$

The exact Cauchy data $E(0,t)$, $H(0,t)$ on the wall $x = 0$ are solutions of (15) and (16): Elimination yields a third degree algebraic equation in $H(0,t)$. For practical purposes, however, we may restrict ourselves to the order of approximation $O(E_L\chi_2)$ as in Sec. 2. Then, from (15) and (16) we find:

$$E(0,t) = 2(K+1)^{-1}E_L\cos(\omega t) - 16\pi[K(K+1)^3]^{-1}(E_L\chi_2)E_L\cos^2(\omega t), \tag{17}$$

$$H(0,t) = (\epsilon_0/\mu_0)^{1/2}\{2K(K+1)^{-1}E_L\cos(\omega t) + 16\pi[K(K+1)^3]^{-1}(E_L\chi_2)E_L\cos^2(\omega t)\},$$

Thus, at the order of approximation stated above, (2) should be replaced by (17); however, the results obtained in Sec. 2 remain unaltered under this substitution.

The second-harmonic reflected wave for $x < 0$ is (approximately) given by:

$$E(x,t) = -8\pi[K(K+1)^3]^{-1}(E_L\chi_2)E_L \times \cos(2\omega t + 4\pi\lambda^{-1}x),$$

$$H(x,t) = 8\pi[K(K+1)^3]^{-1}(E_L\chi_2)(\epsilon_0/\mu_0)^{1/2}E_L \times \cos(2\omega t + 4\pi\lambda^{-1}x), \tag{18}$$

(the 4π factor is due to the definition of the nonlinear optical coefficient as assumed here).

Thus, it remains proved that a (second-) harmonic reflected wave necessarily exists for $x < 0$ if the solution inside the half-space $x \geqslant 0$ is a (nonlinear) progressive simple wave, i.e., if reflection from $x = +\infty$ is excluded. Its (approximate) analytical expression (18) coincides with that obtained by application of the KBM (Krylov–Bogoliubov–Mitropol'skiĭ) perturbative method.[17]

[1]D. Graffi, Atti Accad. Naz. Modena **4**, 9 (1967).
[2]D. Graffi, in *Trends in Applications of Pure Mathematics to Mechanics*, edited by G. Fichera (Pitman, London, 1975).
[3]L. Cesari, Rend. Sem. Mat. Fis. Univ. Milano **45**, 139 (1974).
[4]L. Cesari, Ann. Sc. Norm. Sup. Pisa **4**, 3–4 (1974).
[5]L. Cesari, "The duplication of frequency in laser radiation through nonlinear media," Report, Univ. of Michigan. 1978.
[6]L. Cesari, "Numerical methods for quasilinear hyperbolic systems and applications to a problem of nonlinear optics," to appear.
[7]P. Bassanini, J. Appl. Math. Phys. (ZAMP) **27**, 4, 6 (1976).
[8]P.A. Franken and J.F. Ward, Rev. Mod. Phys. **35**, 1 (1963).
[9]N. Bloembergen, *Nonlinear Optics* (Benjamin, New York, 1977).
[10]S. Singh, *Nonlinear Optical Materials*, C.R.C. Handbook of Lasers (CRC Press, New York, 1971).
[11]P. Bassanini and M.C. Salvatori, "Problemi ai limiti per sistemi iperbolici quasilineari e generazione di armoniche ottiche", to appear in Riv. Mat. Univ. Parma.
[12]P. Bassanini and M.C. Salvatori, "A theorem of existence and uniqueness in nonlinear optics," to appear in Boll. U.M.I.
[13]P.D. Lax, J. Math. Phys. **5**, 611 (1964).
[14]P. Bassanini, Boll. U.M.I. **15-A**, 214 (1978).
[15]P.D. Lax, Commun. Pure Appl. Math. **10**, 537 (1957).
[16]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1965), Vol. II.
[17]D. Graffi, invited lecture, Analytical Problems in Mechanics, CISM (Centre International de Sciences Mécaniques), Udine, 1978.
[18]A. Jeffrey, J. Math. Mech. **17**, 331 (1967).

Solutions of the nonlinear 3-wave equations in three spatial dimensions

H. Cornille

Cen–Saclay, Boite Postale No. 2, 91190 Gif-Sur-Yvette, France
(Received 16 November 1978)

Recently Ablowitz and Haberman have shown that, in three spatial dimensions, the nonlinear 3-wave evolution equation results from the compatibility condition between two well-defined first-order linear differential 3×3 systems having common solutions. We construct inversionlike integral equations (I.E.) associated with both these two linear differential systems so that the solutions of the I.E. embody their compatibility conditions. The scalar kernels of these I.E. depend upon three independent variables in such a way that there exist degenerate kernels confined in the three-dimensional coordinate space. Consequently we exhibit, for the nonlinear 3-wave evolution equations, an infinite number of solutions which, at fixed time, are confined in the three-dimensional coordinate space.

1. INTRODUCTION

Since the last decade there has been a great interest in the explicit construction of solutions of a class of nonlinear partial differential evolution equations (n.l.p.d.e.), which we think are solvable because they represent the compatibility condition between different linear differential systems.[1,2] Within the so-called inversionlike integral equation (I.E.) method, which consists[3] of the construction of a class of potentials and solutions associated with a linear system (without introducing the data), the above n.l.p.d.e. must result from the construction of I.E. which are *common* to these different linear differential systems. Recently another problem appeared as very interesting: namely, the explicit construction of simple really confined solutions of multidimensional n.l.p.d.e. We have previously developed[3,4] a method of constructing I.E. associated with multidimensional linear differential systems. Whereas in two-spatial dimensions the potential reconstructed from the I.E. are free and can be confined in the coordinate space, on the contrary in three or more than three coordinate space[4] the potentials are subject to constraints representing always the compatibility conditions between different linear differential systems. For these dimensional cases, starting from a linear differential system, our I.E. is always associated with at least one other one.

The most simple three-dimensional spatial n.l.p.d.e. is the nonlinear three wave equation (an interesting equation [5a] occurring in plasmas and nonlinear optics) which, from the Ablowitz–Haberman and Zakharov–Shabat[2] works, we know is obtained as the compatibility condition between two 3×3 differential systems. Zakharov [5b] has previously defined a formalism (thereby extending the work in Ref. 2) leading to a class of solutions of the nonlinear three-wave equations in three spatial dimensions. However, among the three coordinates he chooses only one as the variable, whereas the other are parameters. In the present paper, all coordinates are on the same footing, and our formalism is the most general one presented up to now. On the other hand Zakharov does not establish the link between the solutions of the associated linear systems and the solutions of the linear integral equation whose restrictions are the solutions of the resulting nonlinear evolution equation. This link is established here and it may be very important for the complete understanding of the problem. On the other hand, Zakharov does not study the possibility of confined solutions which is a very important part of the present paper.

Later we rederive this compatibility result choosing two adequate linear differential systems. In Sec. 2 we consider a very simple representation of the solutions common to both systems and explicitly construct the common I.E. In our method, the scalar F_j^i kernels of the I.E. associated with a 3×3 first-order differential system must satisfy three linear partial differential equations (l.p.d.e.). However, here both systems *lead to the same set* of three l.p.d.e. Consequently, the F_j^i, which *a priori* depend upon six variables, can be represented as a function of three independent ones. Although different choices of the three remaining variables are possible, we choose them in such a way that degenerate F_j^i kernels can be written down as the product of three arbitrary functions, each of them confining, at finite time, in one spatial direction of the three-dimensional coordinate space. Consequently, we find that the potentials reconstructed from these degenerate kernels of the I.E. are confined in the three-dimensional coordinate space.

In Sec. 3 we start with a general representation of the solutions common to both systems and deduce the corresponding I.E. In this case essentially the same features appear; the I.E. depend upon three independent variables in such a way that degenerate kernels lead to confined potential solutions in the three-dimensional coordinate space. Now we rederive the same nonlinear three waves equations as Ablowitz and Habermann,[2] in a slightly different manner such that:

(i) the intrinsic properties of the n.l.p.d.e. will clearly appear,

 0022-2488/79/081653-14$01.00

(ii) the two associated linear terms continue to have simple common solutions when the potentials are put equal to zero (which is a preliminary convenient step in order to construct common I.E.).

We start with two 3×3 linear differentials systems with four variables x_1, x_2, x_3, x_4 (three of them that we do not specify are associated with the coordinates space and another with the coordinate time),

$$(L^{\pm} - ik\Lambda^{\pm} + R^{\pm})\Psi = 0, \qquad (1\pm)$$

where $\Lambda^{\pm}$ are linked diagonal eigenvalues matrices $\Lambda^{\pm} = (\lambda_i^{\pm}\delta_{ij})$, $R^{\pm}$ are the two 3×3 potentials, $L^{\pm}$ are linked diagonal matrice partial differential operators $L^{\pm} = (l_i^{\pm}\delta_{ij})$, and Ψ is a column vector common to the two systems:

$$R^{+} = \begin{pmatrix} 0 & q_1^2 & 0 \\ 0 & 0 & q_2^3 \\ q_3^1 & 0 & 0 \end{pmatrix}, \quad R^{-} = \begin{pmatrix} 0 & 0 & q_1^3 \\ q_2^1 & 0 & 0 \\ 0 & q_3^2 & 0 \end{pmatrix},$$

$$\Psi = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix},$$

$$l_i^{-} = l_{i+1}^{+}, \quad l_i^{+} = l_i = \sum_{\mu=1}^{4} a_{i\mu}\frac{\partial}{\partial x_\mu}, \quad \lambda_i^{-} = \lambda_{i+1}^{+},$$

where $i = 1,2,3$ ($i+1 = 2,3,1$) are a cyclic permutation. So L^{-} and Λ^{-} are simple cyclic permutations of L^{+} and Λ^{+}.

The compatibility condition is $[L^{+},L^{-}]\Psi = 0$ or $ik(\Lambda^{+}L^{-}\Psi - \Lambda^{-}L^{+}\Psi) + L^{+}(R^{-}\Psi) - L^{-}(R^{+}\Psi) = 0$. Taking into account the relation $L^{+}(R^{-}\Psi) - L^{-}(R^{+}\Psi) = [L^{+}R^{-} - L^{-}R^{+} + R^{-}(ik\Lambda^{-} - R^{-}) - R^{+}(ik\Lambda^{+} - R^{+})]\Psi$, we eliminate Ψ and get $0 \equiv ik[\Lambda^{-}R^{+} - R^{+}\Lambda^{+} - \Lambda^{+}R^{-} + R^{-}\Lambda^{-}] + [L^{+}R^{-} - L^{-}R^{+} + (R^{+})^2 - (R^{-})^2]$. The first bracket is identically zero whereas the second one written with scalar quantities leads to a set of n.l.p.d.e. between the six q_j^i's.

$$\left.\begin{aligned} l_j q_j^{j-1} + q_j^{j+1} q_{j+1}^{j-1} = 0 \\ l_j q_{j-1}^{j} + q_{j-1}^{j+1} q_{j+1}^{j} = 0 \end{aligned}\right\}, \quad l_j = \sum_{1}^{4} a_{j\mu}\frac{\partial}{\partial x_\mu}, \qquad (2)$$

where $j = 1,2,3$ ($j-1 = 3,1,2$) are still a cyclic permutation. If further q_j^{j-1} or q_{j-1}^{j} [or $(q_{j-1}^{j})^{*}$] are proportional, then Eq. (2) can be written with only three different q_j^i's, and we get the nonlinear three-wave evolution equations in a three-dimensional coordinate space. We add two remarks associated with the structures of both Eqs. ($1\pm$) and (2). We remark that the same differential forms $l_j, j = 1,2,3$, appear in both Eq. ($1\pm$) and Eq. (2) and we shall discuss their dimensionality.

(i) *We first discuss the three dimensional coordinate space at fixed time*: From the twelve $\{a_{i\mu}\}$ coefficients of the linear differential forms l_i let us define four vectors

$$\mathbf{a}_\mu = \begin{pmatrix} a_{1\mu} \\ a_{2\mu} \\ a_{3\mu} \end{pmatrix}, \quad \mu = 1,\ldots,4.$$

We require that the l_i be really independent in the three-dimensional coordinate space (which is not specified at this stage). So we assume

$$\det(\mathbf{a}_\mu,\mathbf{a}_\rho,\mathbf{a}_\beta) \neq 0, \quad \mu,\rho,\beta \text{ all different.} \qquad (3)$$

Otherwise, let us, for instance, assume that the time is $t = x_4$ and the spatial coordinates are x_1,x_2,x_3 with $\det(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) = 0$. We define

$$\rho_i = \begin{pmatrix} \rho_{i1} \\ \rho_{i2} \\ \rho_{i3} \end{pmatrix}, \quad i = 1,2,3,$$

$\det(\rho_{ij}) \neq 0$ and $(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)(\boldsymbol{\rho}_i) = 0$ for $i = 1$ and is different of zero for $i = 2,3$. We consider a change of spatial coordinates $z_i = \Sigma_1^3 \rho_{ij} x_j$ and f an arbitrary function of x_μ, $\mu = 1,\ldots,4$, which we write with the new variables z_1,z_2,z_3,t:

$$l_i f = a_{i4}\frac{\partial f}{\partial t} + \sum_{l=2}^{l=3}\left(\sum_{j=1}^{3} a_{ij}\rho_{lj}\right)\frac{\partial f}{\partial z_l}.$$

Consequently, among the spatial derivatives, $\partial f/\partial z_1$ is not present and z_1 would be essentially a parameter. If Eqs. (2) are of this type, they represent in fact a n.l.p.d.e. in a two-dimensional space plus one dimensional time which has been previously solved.[4] At fixed time the confined solutions are confined in a plane and not in a three-dimensional space.

(ii) *We secondly assume always the validity of Eq.* (3) in the following of the paper and discuss the four (space-+ time) dimensional x_μ ($\mu = 1,\ldots,4$) associated with Eqs. ($1\pm$) and (2). We note that for the $\{l_i\}$ or the $\{a_{i\mu}\}$ μ runs from 1 to 4 whereas i runs from 1 to 3. Consequently, we can always, by a change of the (space + time) variables, define new coordinates y_ν, $\nu = 1,\ldots,4$, $y_\nu = \Sigma_1^4 \eta_{\nu\mu}$, $\det(\eta_{\nu\mu}) \neq 0$, $\Sigma_{\mu=1}^{4} a_{i\mu}\eta_{1\mu} = 0$, $i = 1,2,3$ rewrite

$$l_i f(y_1,\ldots,y_4) = \sum_{\nu=2}^{4}\left(\sum_{\mu=1}^{4} a_{i\mu}\eta_{\nu\mu}\right)\frac{\partial f}{\partial y_\nu}$$

in such a way that $\partial f/\partial y_1$ is missing. In that case y_1 is essentially a parameter while y_2,y_3,y_4 with associated partial derivatives, are really variables (linear combination of the space + time coordinates). By this choice of variables Eq. (2) could appear formally similar to a two-spatial dimensional plus one time dimensional problem. However, let us emphasize that the three $(l_i f)$ cannot be reduced (as in i) to a form with only two different partial derivatives for the space and one for the time. Consequently, the main difference in this paper is that the solutions of Eq. (2) at fixed time must exist in a really three-dimensional coordinate space instead of a two-dimensional one. (This point leads to very important differences concerning the possible confinement properties of the solution.) As an illustration of the general results obtained in the paper we write down a very simple example of solutions of Eq. (2). Let us consider a four dimensional-space X_1,X_2,X_3,X_4 defined by the change of variables:

$$\begin{aligned} X_j &= \sum_{1}^{3}\beta_{jl}x_l, \\ X_4 &= \sum_{1}^{3}\beta_{4l}x_l + x_4, \end{aligned} \qquad (4a)$$

where the constants $\{\beta_{jl}\}$ are determined by solving the four systems of 3×3 algebraic linear equations:

$$\begin{cases} l_j X_j = l_{j+1} X_j = 0, & l_{j-1} X_j + a_{j-1,j} \nu_{j-1,j} = 0, \\ l_j X_4 = 0, & j = 1,2,3, \end{cases} \tag{4b}$$

where ν_{jj+1} are constants and $j = 1,2,3$ is still a cyclic permutation. Let us consider

$$\begin{aligned} Dq_j^{j-1} &= -a_{j+1,j-1}\nu_{j+1,j-1} h_j^{j-1}(X_{j-1}) g_j^{j-1}(X_j) l_j^{j-1}(X_4), \\ Dq_{j-1}^{j} &= -a_{j-1,j}\nu_{j-1,j} h_{j+1}^{j}(X_j) g_{j-1}^{j+1}(X_{j-1}) \times A_{j-1,j+1}^{j}(X_{j+1}) l_{j+1}^{j}(X_4) l_{j-1}^{j+1}(X_4), \\ D &= 1 - \prod_{j=1}^{3} A_{j-1,j+1}^{j}(X_{j+1}) l_{j+1}^{j}(X_4), \\ A_{jk}^{i}(X_k) &= \int_0^\infty g_k^i(u + X_k) h_j^k(u + X_k) du, \end{aligned} \tag{5a}$$

where h_j^i, g_j^i, l_j^i are arbitrary functions; then the reader can directly verify with the help of (4b) that Eq. (5a) are solutions of Eq. (2). If further we assume

$$\lim_{|X_i|\to\infty} g_j^i(X_j) = \lim_{|X_i|\to\infty} h_j^i(X_i) = \lim_{|X_4|\to\infty} l_j^i(X_4) = 0, \tag{6}$$

whatever is the variable time chosen among the x_μ's, the q_j^i Eq. (5) solutions of Eq. (2) are confined in the three remaining dimensional coordinate space (outside the values where the Fredholm determinant D can vanish).

The main result of Sec. 2 is the following. First we start with the I.E. (11), where the kernels F_j^i are of the Eq. (4) type and we get the solutions K_j^i. Secondly we consider the restriction $\hat{K}_j^i = K_j^i(y = x)$ and apply the relations (10) leading to the potentials q_j^i's. It follows that the q_j^i's satisfy the n.l.p.d.e. (2). For confined q_j^i we further assume Eq. (6) for the F_j^i.

In Sec. 3 the main result is the following: We start with (11'), F_j^i being of the Eq. (4') type, determine $K_{j\mu}^i$ and the restrictions $\hat{K}_{j\mu}^i = K_{j\mu}^i(y = x_\mu)$, and finally obtain q_j^i [with Eq. (10')] which satisfy Eq. (2). The confined solutions are obtained by assuming Eq. (6').

In Sec. 4 we extend both formalisms of Sec. 2 and 3 in the cases where the representations (8), (8') and I.E. (10), (10') have integration paths $[-\infty, x_\mu]$ and $[x_\mu, \infty]$. This is the most general result in this paper.

2. INVERSIONLIKE FORMALISM FOR A VERY SIMPLE REPRESENTATION OF THE SOLUTIONS OF EQ. (1 ±)

We apply the method developed in previous papers.[3,4] The first step is to find a set $\{\Psi_j^0\}$ of solutions of (1 ±) when $R^{\pm} = 0$. Secondly, we shall write down a set of representations Ψ_j, common solutions of (1 ±), when $R^{\pm} \neq 0$. This will define a set $\{K_j^i\}$ of transform of $\{\Psi_j\}$ with respect to $\{\Psi_j^0\}$. We shall get two sets of n.l.p.d.e. for the $\{K_j^i\}$ and their links with the q_j^i of (1 ±). Thirdly, we shall construct an I.E. such that the solutions satisfy these two sets of n.l.p.d.e. associated with (1 ±). Further we study the properties of the kernels of the I.E., particularly those leading, at fixed time, to the confinement properties, in a three-dimensional coordinate space, of the q_j^i's reconstructed from the I.E. Finally we provide explicit examples where these confinement properties can be directly verified.

A. A set $\{\Psi_j^0\}$ of solutions of Eq. (1 ±) when $R^{\pm} = 0$

We seek $\Psi_j^0 = (u_j^0(x_1,\ldots x_4)\delta_{ij}), j = 1,2,3, \delta_{ii} = 1, \delta_{ij} = 0$ if $i \neq j$, such that $(L^{\pm} - ik\Lambda^{\pm})\Psi_j^0 = 0$. As a third condition we require that

$$u_j^0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

not be a solution of (1 ±) when $R^{\pm} = 0$. Consequently, we introduce three constants $\gamma_i \neq \lambda_i$ and require

$$\begin{aligned} &(l_m - ik\lambda_m)u_j^0 = 0, \quad m = j, j+1, \\ &(l_{j+2} - ik\gamma_{j+2})u_j^0 = 0, \quad \gamma_{j+2} \neq \lambda_{j+2}, \end{aligned} \tag{7a}$$

where $j = 1,2,3$ is still a cyclic permutation ($j + 2 = j - 1$).

For the explicit construction of these u_j^0, let us consider twelve functions $v_{j\mu}^0$ and constants $\lambda_{j\mu}, j = 1,\ldots,3, \mu = 1,\ldots,4$. In order that Eq. (7) can be satisfied, these functions and constants are subject to the following conditions:

$$\begin{aligned} &u_j^0 = \prod_{\mu=1}^{4} v_{j\mu}^0(x_\mu), \quad \left(\frac{\partial}{\partial x} - ik\lambda_{j\mu}\right) v_{j\mu}^0(x) = 0, \\ &\sum_{\mu=1}^{4} a_{m\mu}\lambda_{j\mu} = \lambda_m, \quad m = j, j+1, \\ &\sum_{\mu=1}^{4} a_{j+2\,\mu}\lambda_{j\mu} = \gamma_{j+2} \neq \lambda_{j+2}. \end{aligned} \tag{7b}$$

For j fixed, one $\lambda_{j\mu}$ is given arbitrarily while the three other are obtained by (7b) from the three algebraic equations which can be solved, due to our condition (3) (on the determinants of the three vectors $\mathbf{a}_\eta$).

Although the condition introduced $\gamma_{j+2} \neq \lambda_{j+2}$ is certainly trivial for the reader familiar with the inversion formalism, we would like to briefly explain the reason. The difficulty is more easily understood by considering a scalar first order case in one dimension, $[a\partial/\partial x - ik + q(x)]u(x) = 0$. Let us try to represent u as $u = u_0(x) + \int_x^\infty K(y,x)u_0(y)dy$ with $(a\partial/\partial x - ik)u_0(x) = 0$. Substituting into the equation, we get $a\int_x^\infty u_0(y)(\partial/\partial x + \partial/\partial y)K dy + u_0(x)q(x) = 0$. There is no trivial link between the restriction $K(y = x)$ and $q(x)$, and if as usual we want to satisfy this relation putting to zero both terms, we get $q \equiv 0$ and for the transform K, $(\partial/\partial x + \partial/\partial y)K \equiv 0$. Coming back to our 3×3 system (1 ±), we essentially get the same result: If u_j^0 is solution of the three equations $(l_m - ik\lambda_m)u_j^0 = 0, m = 1,2,3$, then a representation of Ψ in terms of u_j^0 leads to $q_j^i \equiv 0$ and l.p.d.e. for the transforms.

The fact that u_j^0 is not a solution of $l_{j+2} - ik\lambda_{j+2}$ (or equivalently of $\gamma_{j+2} \neq \lambda_{j+2}$) makes it possible for us to introduce new constants $\nu_{j,\mu}$ defined by

$$\left(ik(\gamma_{j+2} - \lambda_{j+2}) + \nu_{j+2,\mu}a_{j+2,\mu}\frac{\partial}{\partial x_\mu}\right)u_j^0 = 0. \tag{7c}$$

In the following we represent Ψ in terms of u_j^0, this procedure defining transforms of Ψ. These constants $\nu_{j\mu}$ represent, in some manner, the strength of the nonlinearity that the transforms have to satisfy. (We notice that the λ_j, γ_j, $\lambda_{j\mu}$, $\gamma_{j\mu}$ are real numbers.)

B. A simple representation of the solutions of (1 ±)

We define three solutions Ψ_j, $j = 1,2,3$, common to both (1 ±):

$$\Psi_j = \left(U_j^0(x_1,x_2,x_3,x_4)\delta_{ij} + \int_{x_j}^{\infty} K_i^j(y;x_1,x_2,x_3,x_4) \times U_j^0(x_j = y)dy \right) \tag{8}$$

$$U_j^0(x_j = y) = V_{jj}^0(y), \quad \left(\prod_{\mu \neq j} V_{j\mu}^0(x_\mu) \right), \quad j = 1,2,3, \quad i = 1,2,3.$$

In this very simple representation x_4 plays a particular role because in Eq. (8) the integration path starts at either x_1,x_2,x_3 but not at x_4. In Sec. 3 we shall consider a general representation symmetric with respect to all the x_μ's. We assume

$$\lim_{y \to +\infty} K_j^i(y,x_1,\ldots,x_4)U_j^0(x_j = y) = 0 \tag{9}$$

and substitute Ψ_j in both systems (1 ±). We get that the $\{K_j^i\}$ must satisfy two well-defined sets of n.l.p.d.e. and the restrictions $\{\hat{K}_j^i = K_j^i(y = x_i)\}$ are linked to the potentials $\{q_j^i\}$.

$$\begin{aligned}
&O_j^{i\,+} K_j^i = a_{j,j+1}\nu_{j,j+1}\hat{K}_j^{j+1}K_{j+1}^i, \\
&O_j^{i\,-} K_j^i = O_{j+1}^{i\,+} K_j^i = a_{j+1,j-1}\nu_{j+1,j-1}\hat{K}_j^{j-1}K_{j-1}^i, \\
&O_j^{i\,+} = l_j + b_{ji}\frac{\partial}{\partial y}, \\
&b_{ji} = a_{ji} \quad \text{except} \quad b_{j,j+1} = a_{j,j+1}(\nu_{j,j+1} + 1), \\
&q_j^{j+1} + a_{j,j+1}\nu_{j,j+1}\hat{K}_j^{j+1} = 0, \\
&q_{j+1}^j + a_{j-1,j}\nu_{j-1,j}\hat{K}_{j+1}^j = 0,
\end{aligned} \tag{10}$$

where $j = 1,2,3$ is still a cyclic permutation. We first remark that if $\gamma_j = \lambda_j$ or $\nu_{j,\mu} = 0$ [or equivalently $(l_{j+2} - ik\lambda_{j+2})U_j^0 = 0$], then the formalism breaks down $q_j^i \equiv 0$ and the nonlinear parts of the n.l.p.d.e. vanish in Eq. (10). Our second remark is that among the set of n.l.p.d.e. for the $\{K_j^i\}$ in Eq. (10) there exists a subset for the $\{\hat{K}_j^i\}$, $i \neq j$,

$$\begin{aligned}
&l_j\hat{K}_j^{j-1} = a_{j,j+1}\nu_{j,j+1}\hat{K}_j^{j+1}\hat{K}_{j+1}^{j-1}, \\
&l_j\hat{K}_{j-1}^j = a_{j,j+1}\nu_{j,j+1}\hat{K}_{j-1}^{j+1}\hat{K}_{j+1}^j,
\end{aligned} \tag{2'}$$

which represents the n.l.p.d.e. (2) written with the $\{\hat{K}_j^i\}$ instead of the $\{q_j^i\}$.

Consequently, if we are able to construct a formalism (I.E.) solving the n.l.p.d.e. (10), then automatically the solutions contain also the solutions of Eq. (2).

C. An inversionlike integral equation (I.E.) associated with Eqs. (8)–(10)

Let us consider the following integral equation:

$$\begin{aligned}
K_j^i(y;x_1,x_2,x_3,x_4) &= \tilde{F}_j^i(y;x_1,x_2,x_3,x_4) \\
&+ \sum_{m=1}^{3} \int_{x_m}^{\infty} F_m^i(s;y;x_1,x_2,x_3,x_4)K_j^m(s;x_1,x_2,x_3,x_4)ds, \\
\tilde{F}_j^i(y;x_1,\ldots,x_4) &= F_j^i(s = x_j;y;x_1,\ldots,x_4).
\end{aligned} \tag{11}$$

We remark that the free terms $\tilde{F}_j^i$ is the restriction when $s = x_j$ of the kernel F_j^i. For each kernel $F_j^i(s;y;x_1,\ldots,x_4)$ we assume the boundary condition

$$\lim_{y \to \infty} F_j^i = 0, \quad \lim_{s \to \infty} F_j^i K_l^j(s;x_1,\ldots,x_4) = 0, \tag{12}$$

and that they satisfy three independent l.p.d.e.

$$\left(l_m + b_{m,i}\frac{\partial}{\partial y} + b_{m,j}\frac{\partial}{\partial s} \right) F_j^i(s;y;x_1,x_2,x_3,x_4) = 0, \quad m = 1,2,3, \tag{13}$$

and we assume, of course, that the solution of Eq. (11) exists and is unique.

Property: If we assume that the $\{F_j^i\}$ satisfy both Eqs. (12) and (13), then the solutions $\{K_j^i\}$ of Eq. (11) satisfy both the two sets of n.l.p.d.e. (10). For the proof given in Appendix A.1 we remark that [due to Eq. (13)] $O_j^{i\pm}\tilde{F}_j^i = 0$ and we apply $O_j^{i\pm}$ to both sides of Eq. (11).

In order to understand more clearly this result let us remark that each kernel F_j^i links three solutions K_m^i ($m = 1,2,3$), via the integral relations $\int F_j^i K_m^j$, to three other ones K_m^j. In order to get an I.E. for quantities like $O_m^{i\,+} K_m^i$, we expect that the F_j^i must satisfy three l.p.d.e. of the type Eq. (12). However, here we want to obtain integral equations for both $O_m^{i\pm} K_m^i$ and we expect *a priori* more than three constraints for the F_j^i. What seems miraculous is the fact that *the same set* Eq. (13) of l.p.d.e. works in both cases. The deep meaning must be understood in connection with our previous work.[4] In the more than two dimensional case, the I.E. of Eq. (11) type built for a first order linear differential system is in fact associated also with "ghost" linear first order systems [for a 3×3 system like Eq. (1 +) or Eq. (1 −) with only one other "ghost" linear system]. This means that the same set of l.p.d.e. for the $\{F_j^i\}$ must work for two associated linear systems. Here we have managed the formalism in such a way that the "ghost" system associated with Eq. (1 +) [or Eq. (1 −)] is just Eq. (1 −) [or Eq. (1 +)].

D. Properties of the kernels F_j^i

The F_j^i depend upon six variables s,y,x_1,x_2,x_3,x_4, subject to three first order l.p.d.e. (13). Moreover, to each of these six variables, corresponds a first order partial derivative. This means that by a linear combination of these variables we expect to define three independent new variables. In order to get more easily degenerate kernels, we require that one new variable constains s, another y, and the last one neither s nor y. Let us define four constant vectors

$$\beta_\mu = \begin{pmatrix} \beta_{\mu 1} \\ \beta_{\mu 2} \\ \beta_{\mu 3} \end{pmatrix}, \quad \mu = 1,\ldots,4$$

and consider the change of coordinates

$$X_j = \sum_1^3 \beta_{jl} x_l, \quad j = 1,2,3, \quad X_4 = x_4 + \sum_1^3 \beta_{4l} x_l.$$

Assuming that the F_j^i are of the form

$$F_j^i(s - x_j + X_j; y - x_i + X_i; X_4) \tag{4'}$$

and substituting into the three l.p.d.e. (13) (where m is either $j, j \pm 1$; $i, i \pm 1$ or 1,2,3), we get that the four sets $\{\beta_{\mu l}\}$ (μ fixed, $l = 1,2,3$) are solutions of four algebraic 3×3 linear systems:

$$\sum_{l=1}^3 a_{ml}\beta_{jl} = 0, \quad m = j, j+1,$$

$$\sum_{l=1}^3 a_{j-1,l}\beta_{jl} + a_{j-1}v_{j-1,j} = 0, \quad j = 1,2,3,$$

$$\sum_{l=1}^3 a_{ml}\beta_{4l} + a_{m,4} = 0, \quad m = 1,2,3. \tag{4b'}$$

(4b') are nothing else but the relations (4b) of the Introduction written down here. Due to the assumption (3) for $\det(\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3)$, the solutions of (4b') exist. Let us now show that the q_j^i reconstructed from Eq. (11) depend in fact upon only the four X_μ ($\mu = 1,...,4$). Let us define $G_j^i(z) = K_j^i(y = x_i + z)$, and, substituting Eq. (4) for the F_j^i kernels, we can rewrite Eq. (11):

$$G_j^i(z) = F_j^i(X_j; z + X_i; X_4) + \sum_m \int_0^\infty F_m^i(u + X_m; z + X_i; X_4) \times G_j^m(u)du. \tag{11'}$$

It follows that $G_j^i(z)$ depends upon five variables z, X_μ, $\mu = 1,...,4$, and $G_j^i(z = 0) = \hat{K}_j^i$ (or q_j^i) depends upon X_μ, $\mu = 1,...,4$. Consequently, the properties of the solutions of Eq. (11') can be studied in this new four-dimensional space X_μ. In Appendix A.2 the following properties are established:

(i) $\{X_\mu\}, \mu = 1,...,4$, determines a four-dimensional space.

(ii) If the time is $t = x_4$, then, at fixed time, x_1, x_2, x_3 as well as $X_j, X_i, X_4 - t$ ($i \neq j$, i and j being either 1,2,3) determine a three-dimensional coordinate space.

(iii) If the time is $t \neq x_4$, for instance $t = x_3$, then, at fixed time, x_1, x_2, x_4 as well as $X_i - t\beta_{i3}, X_j - t\beta_{j3}, X_4 - t\beta_{43}$ determine a three-dimensional coordinate space.

From Eq. (11) or Eq. (11'), we remark that for $i \neq j$

$$\hat{K}_j^i = \tilde{F}_j^i(y = x_i) + \text{other terms}$$

$$\tilde{F}_j^i(y = x_i) = F_j^i(X_j; X_i; X_4). \tag{14}$$

If we investigate at fixed time the possible confinement properties of the solutions in the three-dimensional coordinate space, we must at least look at the first term $\tilde{F}_j^i(y = x_i)$. We note that degenerate F_j^i kernels of the type

$$F_j^i = \sum_{m=1}^{m_0} g_{jm}^i(s - x_j + X_j)h_{jm}^i(y - x_i + X_i)l_{jm}^i(X_4)$$

lead to $\tilde{F}_j^i(y = x_i) = \Sigma_m g_{jm}^i(X_j)h_{jm}^i(X_i)l_{jm}^i(X_4)$. We consider $i \neq j$ and take for $g_{jm}^i, h_{jm}^i, l_{jm}^i$ functions of the Eq. (6) type going to zero when their variables are going to $\pm\infty$. From the above results it follows that, in the (ii) case ($t = x_4$), either in the x_1, x_2, x_3 coordinate space or equivalently in the $X_i, X_j, X_4 - t$ space, the $|\tilde{F}_j^i(y = x_i)|$ as well as the $|F_j^i|$ of Eq. (11) are confined. Similarly in the (iii) case ($t = x_3$), either in the x_1, x_2, x_4 coordinate space or equivalently in the $X_i - t\beta_{i3}, X_j - t\beta_{j3}, X_4 - t\beta_{43}$ space, the $|\tilde{F}_j^i(y = x_i)|$ and the $|F_j^i|$ are confined.

Let us define $\mathcal{K}(y; x_1,...,x_4) = (K_j^i)$, $\tilde{\mathcal{F}} = (\tilde{F}_j^i)$, $\mathcal{F} = (F_j^i\theta(s - x_j))$, then Eq. (11) can be written in a matrix form:

$$\mathcal{K}(y; x_1...x_4) = \tilde{\mathcal{F}}(y; x_1, x_2, x_3, x_4) + \int_{-\infty}^{+\infty} \mathcal{F}(s; y; x_1,...,x_4)\mathcal{K}(s; x_1,...,x_4).$$

In the following we continue our investigations of the confinement properties of the solutions of Eq. (11) in the three-dimensional space, at fixed time, in cases of degenerate F_j^i kernels where these solutions can be written down explicitly. We do not distinguish between $t = x_4$ and $t \neq x_4$ and, always put $F_i^i = 0$ and take into account the properties obtained in (ii) and (iii).

E. Simple examples

We assume that $F_i^i = 0$ and for $i \neq j$ the most simple degenerate kernels

$$F_j^i(s - x_j + X_j; y - x_i + X_i; X_4) = g_j^i(s - x_j + X_j)h_j^i(y - x_i + X_i)l_j^i(X_4),$$

g_j^i, h_j^i, l_j^i of the Eq. (6) type. (15)

(1) We first consider

$$\mathcal{F} = \begin{pmatrix} 0 & F_2^1\theta(s - x_2) & 0 \\ 0 & 0 & F_3^2\theta(s - x_3) \\ F_1^3\theta(s - x_1) & 0 & 0 \end{pmatrix}. \tag{5b}$$

The solutions are written down in the Introduction, Eq. (5a). $|A_{jk}^i(X_k)|$ are bounded and the Fredholm determinant D is bounded (in all discussion in this paper we disregard the cases when the Fredholm determinant vanishes). In the numerator of the solutions of Eq. (5) we see that there always exist a product of three functions of the Eq. (6) type, each of them confining in a direction of the three dimensional coordinate space. Consequently, the reconstructed potentials q_j^i are confined in the three-dimensional coordinate space.

(ii) Secondly we consider always F_j^i like Eq. (15) and

$$\mathcal{F} = \begin{pmatrix} 0 & F_2^1\theta(s - x_2) & F_3^1\theta(s - x_3) \\ F_1^2\theta(s - x_1) & 0 & 0 \\ F_1^3\theta(s - x_1) & 0 & 0 \end{pmatrix}, \tag{16}$$

the corresponding solutions are written down in Table Ia. Here also D is bounded and $D\hat{K}_j^i$ is always a linear combination of terms such that each of them is a product of functions, confining in three different spatial dimensions, multiplied by bounded functions. Consequently, for finite time, the solutions are confined in a three-dimensional space.

TABLE I.a.

$$D\hat{K}^2_1 = h^2_1 l^2_1 [g^2_1 + A^1_{13} l^1_3 l^3_1 (g^3_1 A^2_{31} - g^2_1 A^3_{31})]\, D\hat{K}^2_3 = h^2_1 g^1_3 l^1_3 l^2_1 A^2_{31}$$
$$D\hat{K}^1_3 = g^1_3 l^1_3 [h^1_3 + A^1_{12} l^1_2 l^2_1 (h^1_2 A^2_{31} - h^1_3 A^2_{31})]$$
$$D = (1 - A^1_{12} A^2_{21} l^1_2 l^2_1)(1 - A^3_{31} A^1_{13} l^3_1 l^1_3) - A^1_{13} A^1_{12} A^2_{31} A^3_{21} l^1_3 l^3_1 l^1_2 l^2_1$$
$$\hat{K}^2_1 \leftrightarrow \hat{K}^3_1 \text{ if } 2 \leftrightarrow 3, \quad \hat{K}^1_3 \leftrightarrow \hat{K}^1_2 \text{ if } 3 \leftrightarrow 2, \quad \hat{K}^2_3 \leftrightarrow \hat{K}^3_2 \text{ if } 3 \leftrightarrow 2$$
$$h^i_j = h^i_j(X_i) g^i_j = g^i_j(X_j) l^i_j = l^i_j(X_4) A^i_{jk} = \int_0^\infty g^i_k(u + X_k) h^k_j(u + X_k)\, du$$

TABLE I.b.

$$DB_1 = -a_{31}\nu_{31}\sigma^1_3 l^1_3 (h^3_1)^* \left[h^1_3 + |l^1_2|^2 \sigma^1_2 \sigma^2_1 \int |h^2_1|^2 \left(h^1_2 \int h^{1*}_2 h^1_3 - h^1_3 \int |h^1_2|^2 \right) \right]$$
$$DB_2 = -a_{31}\nu_{31}\sigma^1_2 l^1_2 (h^2_1)^* \left[h^1_2 + |l^1_3|^2 \sigma^1_3 \sigma^3_1 \int |h^3_1|^2 \left(h^1_3 \int h^1_2 h^{1*}_3 - h^1_2 \int |h^1_3|^2 \right) \right]$$
$$D = 1 - \sigma^1_2 \sigma^2_1 |l^1_2|^2 \int |h^2_1|^2 \int |h^1_2|^2 - \sigma^1_3 \sigma^3_1 |l^1_3|^2 \int |h^3_1|^2 \int |h^1_3|^2$$
$$+ \sigma^1_2 \sigma^2_1 \sigma^1_3 \sigma^3_1 |l^1_2 l^1_3|^2 \int |h^2_1|^2 \int |h^3_1|^2 \left[\int |h^1_2|^2 \int |h^1_3|^2 - \left| \int h^{1*}_2 h^1_3 \right|^2 \right]$$
$$h^i_j = h^i_j(X_i), \quad l^i_j = l^i_j(X_4), \quad \int |h^i_j|^2 = \int_0^\infty |h^i_j(u + X_i)|^2 du,$$
$$\int h^{i*}_j h^i_k = (h^i_j(u + X_i))^* h^i_k(u + X_i) du.$$

TABLE I.c.

$$D\hat{K}^3_{2\mu_i} = g^1_{2\mu_i} h^3_{1\mu_i} l^1_{2\mu_i} A^3_{21\mu_i\mu_i} \quad D\hat{K}^3_{1\mu_i} = h^3_{1\mu_i} [g^3_{1\mu_i} l^3_{1\mu_i} + A^1_{32\mu_i} A^2_{13\mu_i} (g^3_{1\mu_i} l^3_{1\mu_i} A^3_{1\mu_i\mu_i} - g^3_{1\mu_i} l^3_{1\mu_i} A^3_{21\mu_i\mu_i})], \quad j = 3,4, \; l = 3,4, \; j \neq l$$
$$D\hat{K}^i_{1,1\mu_i} = g^i_{1,1\mu_i} l^i_{1,1\mu_i} h^i_{1,1\mu_i}, \quad i = 1,2$$
$$D\hat{K}^1_{3\mu_i} = g^2_{3\mu_i} l^2_{3\mu_i} h^1_{2\mu_i} A^1_{32\mu_i\mu_i}, \quad D\hat{K}^1_{2\mu_i} = h^2_{3\mu_i} \sum_{j=3}^{4} g^3_{1\mu_i} l^3_{1\mu_i} A^2_{13\mu_i\mu_i}$$
$$A^i_{jk\mu\rho} = l^i_{k\mu}(X^\mu_4) \int_0^\infty g^{i\rho}_{k\mu}(\gamma^\rho_k u + X^\mu_k) h^k_{j\rho}(u + X^\rho_k) g^i_{j\mu} = g^i_{j\mu}(X^\mu_j)$$
$$h^i_{j\mu} = h^i_{j\mu}(X^\mu_i) l^i_{j\mu} = l^i_{j\mu}(X^\mu_4)$$

F. A general degenerate case

Let us assume $F^i_i = 0$ and for $i \neq j$

$$F^i_j = \sum_{m=1}^{m_0} g^i_{jm}(s - x_j + X_j) h^i_{jm}(y - x_i + X_i) l^i_{j,m}(X_4),$$

$g^i_{j,m}, h^i_{jm} l^i_{jm}$ of the Eq. (6) type, m_0 arbitrary finite integer. (17)

Even if the solutions are too complicated to be written in closed form [like (5a) or Table I.a when $m_0 = 1$] we can show the confinement properties. For kernels like (17) the solution of Eq. (11) is reduced to the resolution of algebraic systems and the general solution is established in Appendix A.3. We consider fixed time values (t being either x_4 or any $x_j, j = 1,2,3$) and study the confinement properties in the remaining three-dimensional space [using the results of subsection D, (ii) and (iii)]. We first have to solve a linear algebraic system for the quantities $\{A^j_{jip}, A^k_{jip}$, and $i \leftrightarrow k\}$, $p = 1,\ldots,m_0$ where $A^i_{jkp} = \int g^i_{kp} K^k_j$. We remark (see Appendix A.3) that these quantities have the structure (always excluding the zeros of the Fredholm determinant)
$\Sigma g^i_{jm}(X_j) l^i_{jm}(X_4) \times$ bounded functions $+ \Sigma g^k_{jm}(X_j) l^k_{jm}(X_4) \times$ bounded functions. Substituting these quantities into the expression of $\hat{K}^i_j$ (written down in Appendix A.3), then $\hat{K}^i_j$ is always a linear combination of crucial terms multiplied by bounded functions, these crucial terms being the product of three functions (at least) confining into three different dimensions of the three-dimensional space (at fixed time).

G. Confined solutions, at fixed time, of the three waves n.l.p.d.e.in a three-dimensional space

In order to get this n.l.p.d.e. between three q^i_j's of Eq. (2) we must link q^i_j and $(q^i_j)^*$ or F^i_j and $(F^i_j)^*$. Doing this, we do not alter the confinement properties previously studied.

We consider the simple example given in Eqs. (15)–(16), assume

$$g^i_j(X_j) = \sigma^i_j (h^j_i(X_j))^*, \quad l^i_j(X_4) = (l^i_j(X_4))^*, \quad \sigma^i_j \text{ real constants}, \tag{18}$$

and substitute into the solution written down in Table I.a. Let us define $q^1_3 = B_1$, $q^1_2 = B_2$, $q^2_3 = B_3$, then (B_1, B_2, B_3) satisfy the n.l.p.d.e.

$$l_1 B_1 + B_2 B_3 = 0, \quad l_j = \sum_1^4 a_{j\mu} \frac{\partial}{\partial x_\mu},$$
$$l_2 B_2 + B_1 B^*_3 a_{23}\nu_{23}\sigma^3_1\sigma^1_2 (a_{12}\nu_{12}\sigma^2_1\sigma^1_3)^{-1} = 0, \tag{19}$$
$$l_3 B_3 + B_1 B^*_2 a_{12}\nu_{12}\sigma^2_1 (a_{31}\nu_{31}\sigma^1_2)^{-1} = 0.$$

Taking into account Eq. (18) in Table I.a, we eliminate the g^i_j for the expression of the B_i's. Finally we have six independent arbitrary functions $h^1_2, h^1_3, h^2_1, h^3_1, l^1_2, l^1_3$ that we take of the Eq. (6) type. The solutions are written down in Table I.b. If one of the x_μ variables is a fixed time, then the B_i are confined to the three-dimensional space of the remaining x_μ coordinates (outside the valued where D, the Fredholm determinent, is zero). Applying the Schwarz inequality to the third term of D and if $\sigma^1_2\sigma^2_1 < 0$, $\sigma^1_3\sigma^3_1 < 0$, we note that in this case $D \neq 0$ whereas in other cases D can be zero.

This example quoted in Table I.b is an extension of a previously given example in two spatial dimensions.[4] Let us formally put all l^i_j equal to constants; then x_4 which is present only in $l^i_j(X_4)$ disappears in the solutions. The new B_i's still satisfy Eq. (19) with $l_j B_j$ having partial derivatives with respect to only three variables. If further one of the remaining x_1, x_2, x_3 variables is chosen as the time, then our solutions are still confined but now in a two-dimensional space at fixed time. We verify in this simple case that the solutions still exist in lower dimensions. Finally let us remark that if it is easy (at fixed time) to go from an example in a three-dimensional space to its reduction in a two-dimensional one, the converse is in general not trivial.

3. INVERSIONLIKE FORMALISM FOR A GENERAL REPRESENTATION OF THE SOLUTIONS OF EQ. (1 ±)

We consider a general representation where x_4 does not play a particular role compared to the other x_μ.

A. $U_j^0;\Psi_j^0 = (U_j^0\delta_{ij})$

We start with the same U_j^0 defined in the previous section, Eqs. (7a), (7b), (7c) and consider the same $\Psi_j^0 = (U_j^0\delta_{ij})$.

B. A general representation for three independent solutions of Eq. (1 ±)

For each j value (1,2,3) fixed let us define I_j as either the set $\{1,2,3,4\}$ or any nonempty subset of these four integers. Let us consider

$$\Psi_j = \Big(U_j^0(x_1,x_2,x_3,x_4)\delta_{ij} + \sum_{\mu\in I_j} \int_{x_\mu} K^j_{i\mu}(y;x_1,\ldots,x_4)\, U^0_{j\mu}(y)dy \Big), \tag{8'}$$

$$U^0_{j\mu}(y) = v^0_{j\mu}(y) \prod_{n\neq\mu} v^0_{jn}(x_\eta), \quad U^0_{j\mu}(y = x_\mu) = U^0_j, \quad i = 1,2,3,\ j = 1,2,3.$$

This representation generalizes the previous one, Eq. (8), where $I_j = \{j\}$. If I_j has one element, then we have four possibilities: either 1, or 2, or 3, or 4. If I_j has two elements, we have six possibilities; if I_j has three elements, we have four possibilities and only one choice if I_j is $\{1,2,3,4\}$. So for j fixed we have fifteen different possible representations of Ψ_j. This means that for the three solutions Ψ_1,Ψ_2,Ψ_3 we have $(15)^3$ different representations of the solutions of (1 ±). Our aim is to find the I.E. corresponding to these $(15)^3$ different representations of (1 ±). We notice that if all the x_μ are equal, then all these possible representations in Eq. (8') coincide. Let us assume

$$\lim_{y\to\infty} K^j_{i\mu}(y;x_1,\ldots,x_4)U^0_{j\mu}(y) = 0,$$

define $\hat{K}_{j\mu} = K^i_{j\mu}$ $(y = x_\mu)$, and substitute Eq. (8') into Eq. (1 ±). We get two sets of n.l.p.d.e. which generalize Eq. (10):

$$O^{i+}_{j\mu}K^i_{j\mu} + q^{j+1}_j K^i_{j+1\mu} = 0,$$
$$O^{i+}_{j\mu}K^i_{j\mu} - O^{i+}_{j+1,\mu}K^i_{j\mu} = -q^{j-1}_j K^i_{j-1}, \quad \mu\in I_i,$$
$$O^{i+}_{j\mu} = l_j + b^i_{j\mu}\frac{\partial}{\partial y}, \tag{10'}$$
$$b^i_{j\mu} = a_{j\mu} \quad \text{except} \quad b^i_{i-1,\mu} = a_{i-1,\mu}(\nu_{i-1,\mu}+1), \quad \mu\in I_i,$$
$$q^{j+1}_j + \sum_{\mu\in I_{j+1}} a_{j\mu}\nu_{j\mu}\hat{K}^{j+1}_{j\mu} = 0,$$
$$q^j_{j+1} + \sum_{\mu\in I_j} a_{j-1,\mu}\nu_{j-1,\mu}\hat{K}^j_{j+1,\mu} = 0,$$

where $j = 1,2,3$ is still a cyclic permutation. We still remark that if $\{\nu_{j\mu} = 0\}$, then $q^i_j \equiv 0$ and the $K^i_{j\mu}$ satisfy l.p.d.e. We first verify that if $I_i = \{i\}$ or $K^i_{j\mu}\equiv 0$ except $K^i_{ji}\equiv K^i_j$, then Eq. (10') reduces to Eq. (10). For instance, for the system (1 +), we get $q^{j+1}_j + a_{jj+1}\nu_{jj+1}\hat{K}^{j+1}_j = 0$ and substituting this q^{j+1}_j relation in Eq. (10'), we find the first relation of Eq. (10). We can similarly identify (10) and (10') for the system (1 −).

Secondly from Eq. (10') we get a subset of n.l.p.d.e. for the $\{\hat{K}^i_{j\mu}\}$ similar to (2').

$$l_j\hat{K}^{j-1}_{j\mu} + q^{j+1}_j\hat{K}^{j-1}_{j+1,\mu} = 0, \quad \mu\in I_{j-1},$$
$$l_j\hat{K}^j_{j-1\mu} + q^{j+1}_{j-1}\hat{K}^j_{j+1,\mu} = 0, \quad \mu\in I_j. \tag{2''}$$

Multiplying the first Eq. (2'') relation by $a_{j+1,\mu}\nu_{j+1,\mu}$ and the second one by $a_{j-1,\mu}\nu_{j-1,\mu}$ and finally summing over all the corresponding μ values both relations, we obtain the n.l.p.d.e. (2) [taking into account the links between the $\{q^i_j\}$ and $\{\hat{K}^i_{j\mu}\}$ given by Eq. (10')]. Consequently, if we construct an I.E. such that their solutions verify the n.l.p.d.e. (10'), then we obtain the solutions of Eq. (2).

C. An I.E. associated with (8')–(10')

Let us consider:

$$K^i_{j\mu}(y;x_1,\ldots,x_4) = \widetilde{F}^i_{j\mu}(y;x_1,\ldots,x_4) + \sum_{m=1}^{3}\sum_{\rho\in I_m}\int_{x_\rho} F^i_{m\mu\rho}(s;y;x_1,\ldots,x_4)K^m_{j\rho}(s;x_1,\ldots,x_4)ds, \tag{11'}$$
$$\widetilde{F}^i_{j\mu} = F^i_{j\mu\eta}(s = x_\eta;y;x_1,\ldots,x_4), \mu\in I_i,\ \eta\in I_j.$$

We remark that the free term $\widetilde{F}^i_{j\mu}$ is the restriction when $s = x_\rho$ or the kernels $F^i_{j\mu\rho}$ $(\rho\in I_j)$. If (11') is written in a matrix form, then the matrix kernel of the I.E. is a 12×12 matrix if $I_j = \{1,2,3,4\}$ for all $j = 1,2,3$, instead of our previous 3×3 matrix in the $I_j = \{j\}$ or Eq. (11) case. It is clear, of course, that all other intermediate cases are possible depending upon different choices of the three I_j. So we see that Eqs. (11') recover a great number of different possibilities; nevertheless, general results can still be obtained. For each kernel $F^i_{j\mu\rho}$ we assume the boundary conditions

$$\lim_{y\to\infty} F^i_{j\mu\rho} = 0, \quad \lim_{s\to\infty} F^i_{j\mu\rho}K^j_{l\rho} = 0 \tag{12'}$$

and that they satisfy three independent l.p.d.e.

$$\Big(l_m + b^i_{m\mu}\frac{\partial}{\partial y} + b^j_{m\rho}\frac{\partial}{\partial s}\Big)F^i_{j\mu\rho}(s;y;x_1,\ldots,x_4) = 0, \quad m = 1,2,3, \tag{13'}$$

and we assume that the solution of (11') exists and is unique.

Property: If $\{F^i_{j\mu\rho}\}$ satisfy (12'), (13'), then the $\{K^i_{j\mu}\}$ solutions of (11') verify the n.l.p.d.e. (10') (see Appendix A'.1 for the proof).

D. Properties of $F^i_{j\mu\rho}$

These $F^i_{j\mu\rho}$ kernels have six variables s,y,x_1,x_2,x_3,x_4, subject to three l.p.d.e. (13'), and consequently we want to construct three new variables, linear combinations of the old ones. At this stage this is the same problem as in Sec. 2; however, here we have another constraint, namely, that the restriction $s = x_\rho$ of the $F^i_{j\mu\rho}$ kernels $\rho\in I_j$ lead to the same kernel $\widetilde{F}^i_{j\mu}$. However, a feature common to Eqs. (13), (13') is that we have *three* l.p.d.e. ($m = 1,2,3$) and four coordinates x_η, $\eta = 1,\ldots,4$ [this feature is already present in the n.l.p.d.e. (2), as was discussed in the introduction]. This means than one on the x_η must play a particular role, and due to the representation (8) we choose x_4 in Sec. 2. Here in Eq. (8') no x_ν play a particular role, and we choose x_μ for the kernels

$F^i_{j\mu\rho}$ ($\rho \in I_j$). We define new coordinates X^μ_η ($\eta = 1,...,4$) and try $F^i_{j\mu\rho}$ to be of the form

$$F^i_{j\mu\rho}(\gamma^\rho_{j\mu}(s - x_\rho) + X^\mu_j; y - x_\mu + X^\mu_i; X^\mu_4), \quad \mu \in I_i,\ \rho \in I_j,$$
$$X^\mu_j = \sum_{n \neq \mu} \beta^\mu_{j\eta} x_\eta, \quad j = 1,2,3,$$
$$X^\mu_4 = \sum_{n \neq \mu} \beta^\mu_{4\eta} x_\eta + x_\mu, \quad \gamma^\mu_{j\mu} = 1, \tag{4'}$$
$$\gamma^\rho_{j\mu} = a_{j-1,\mu} v_{j-1,\mu} / a_{j-1,\rho} v_{j-1,\rho}.$$

Substituting (4') into Eq. (13'), we find that the four $\{\beta^\mu_{\alpha\eta}\}$, $\alpha = 1,...,4$, are the solutions of 3×3 linear algebraic systems:

$$j = 1,2,3: \sum_{\eta \neq \mu} a_{m\eta} \beta^\mu_{j\eta} = 0, m = j, j+1,$$
$$\sum_{\eta \neq \mu} a_{j-1,\eta} \beta^\mu_{j\eta} + a_{j-1,\eta} v_{j-1,\mu} = 0; \tag{4b''}$$
$$\sum_{\eta \neq \mu} a_{m\eta} \beta^\mu_{4\eta} + a_{m\mu} = 0, \quad m = 1,2,3,$$

whose solutions always exist due to the assumption (3). The q^i_j's reconstructed from Eq. (11') depend upon the (X^μ_η), $\nu = 1,...,4$. Let us define $G^i_{j\mu}(z) = K^i_{j\mu}(y = x_\mu + z)$ and substituting Eq. (4b") into Eq. (11'); we get

$$G^i_{j\mu}(z) = F^i_{j\mu}(X_j; z + X^\mu_i, X^\mu_4) + \sum_m \sum_{\rho \in I_m} \int_0^\infty F^i_{\eta\mu\rho}(\gamma^\rho_{j\mu} u + X^\mu_j; z + X^\mu_i, X^\mu_4) G^m_{j\rho}(u) du. \tag{11''}$$

It follows that $\hat{K}^i_{j\mu} = G^i_{j\mu}(z = 0)$ depend upon the sets (X^μ_η), $\eta = 1,...,4$. In Appendix A'.2a we have studied for μ fixed, the dimensional properties of these new variables and we get:

(i) At μ fixed, (X^μ_η), $\eta = 1,...,4$, determine a four-dimensional space.

(ii) If the time is $t = x_\mu$, then either x_{η_i} ($\eta_i \neq \mu$, $i = 1,2,3$) or $X^\mu_j, X^\mu_i, X^\mu_4 - t$ determine a three-dimensional coordinate space (μ being fixed).

(iii) If the time is $t \neq x_\mu$, for instance $t = x_{\eta_i}$ then either $x_\mu, x_{\eta_1}, x_{\eta_2}$ or $X^\mu_i - t\beta^\mu_{i\eta_1}, X^\mu_j - t\beta^\mu_{j\eta_1}, X_4 - t\beta^\mu_{4\eta_1}$ ($i \neq j$, $i \leqslant 3$, $j \leqslant 3$, $\eta_l \neq \mu$) determine a three-dimensional coordinate space (μ being fixed).

From Eq. (11') we get for $i \neq j$

$$\hat{K}^i_{j\mu} = F^i_{j\mu}(X^\mu_j; X^\mu_i; X^\mu_4) + \text{other terms.} \tag{14'}$$

For the confinement properties of the reconstructed potentials we look at the first term of the expansion. We consider degenerated kernels

$$F^i_{j\mu\rho} = \sum_m g^{i\rho}_{j\mu m}(\gamma^\rho_{j\mu}(s - x_\rho) + X^\mu_j) h^i_{j\mu m}(y - x_\mu + X^\mu_i) l^i_{j\mu m}(X_4), \quad \mu \in I_i,\ \rho \in I_j, \tag{6'}$$

$g^{i\rho}_{j\mu m}(s = x_\rho) = g^i_{j\mu m}(X^\mu_j)$, $h^i_{j\mu m}, l^i_{j\mu m}, g^{i\rho}_{j\mu m}$ being of the Eq. (6) type, going to zero when their arguments are going to infinity. At fixed time, when t equals one of the x_ν's values, applying the results (ii) or (iii) we see that $|F^i_{j\mu}|$ as well as $|F^i_{j\mu\rho}|$ are confined in three-dimensional space constructed with the three other x_ν's values. So the first term of the expansion (14') as well as the kernels of Eq. (11') are confined in this three-dimensional coordinate space. Now we would like to verify explicitly, as in Sec. 2, that the whole solution $\hat{K}^i_{j\mu}$ and q^i_j are confined in this three-dimensional space. However, with kernels Eq. (6'), appear (for the whole solutions) functions of variables $X^{\mu 1}_i$, $X^{\mu 2}_j$ and $X^{\mu 1}_4$ corresponding to mixed μ values and $i \neq j$. The study is done in Appendix A'.2b and we get: (i) If $t = x_{\mu_1}$, then $X^{\mu_1}_i, X^{\mu_2}_j - \beta^{\mu_2}_{j\mu_1} t, X^{\mu_1}_4 - t$ determine a three-dimensional space; (ii) if $t = x_{\mu_2}$, then $X^{\mu_1}_i - \beta^{\mu_1}_{i\mu_2} t, X^{\mu_2}_j, X^{\mu_1}_4 - \beta^{\mu_1}_{4\mu_2} t$ determine a three-dimensional space; (iii) if $t = x_\varphi$, $\varphi \neq \mu_1$, $\varphi \neq \mu_2$, then $X^{\mu_1}_i - \beta^{\mu_1}_{i\varphi} t$, $X^{\mu_2}_j - \beta^{\mu_2}_{j\varphi} t, X^{\mu_1}_4 - \beta^{\mu_1}_{4\varphi} t$ determine a three-dimensional space.

Similarly to the previous section, Eq. (11') can be written in a matrix form. If n is the number of elements of $I_1 \cup I_2 \cup I_3$, then $3 \leqslant n \leqslant 12$ and the matrix $(F^i_{j\mu\rho})$ is an $n \times n$ matrix whereas $(K^i_{j\mu})$ and $(\tilde{F}^i_{j\mu})$ are $n \times 3$ matrices. This emphasizes the great number of cases that the formalism of this section recovers.

E. Simple example

We assume $F^i_{i\mu\rho} \equiv 0$ and for $i \neq j$ the most simple degenerate kernels:

$$F^i_{j\mu\rho} = g^{i\rho}_{j\mu}(\gamma^\rho_{j\mu}(s - x_\rho) + X^\mu_j) h^i_{j\mu}(y - x_\mu + X^\mu_i) l^i_{j\mu}(X^\mu_4),$$
$$g^{i\rho}_{j\mu}(s = x_\mu) = g^i_{j\mu}(X^\mu_j), \quad g^{i\rho}_{j\mu}, h^i_{j\mu}, l^i_{j\mu} \text{ of the Eq. (6) type}, \quad \mu \in I_i,\ \rho \in I_j. \tag{15'}$$

We consider a very simple case $I_1 = \{\mu_1\}$, $I_2 = \{\mu_2\}$, $I_3 = \{\mu_3, \mu_4\}$, where μ_1, μ_4 can take any of the integer values 1,2,3,4. Then $(F^i_{j\mu\rho})$ is a 4×4 matrix whereas $(K^i_{j\mu})$ and $(\tilde{F}^i_{j\mu})$ are 4×3 matrices. For simplicity we further put some of the kernels, $i \neq j$, equal to zero and retain only for $(F^i_{j\mu\rho})$:

$$\begin{pmatrix} 0 & F^1_{2\mu_1\mu_2}\theta(s - x_{\mu_2}) & 0 & 0 \\ 0 & 0 & F^2_{2\mu_2\mu_3}\theta(s - x_{\mu_3}) & F^2_{3\mu_3\mu_4}\theta(s - x_{\mu_4}) \\ F^3_{1\mu_3\mu_1}\theta(s - x_{\mu_1}) & 0 & 0 & 0 \\ F^3_{1\mu_4\mu_1}\theta(s - x_{\mu_1}) & 0 & 0 & 0 \end{pmatrix}. \tag{5b'}$$

The solutions for the corresponding $\hat{K}^i_{j\mu}$ are written down in Table I.c and the q^i_j's are easily obtained with the relation (10'). Using the results of Appendix A'.2a, A'.2b (recalled above), it is easy to verify that, for any choice of the time variable (among the four x_μ's), the q^i_j's are confined to the three remaining coordinate spaces. The criteria, at fixed time, are the same as in the

previous section. We have linear combination of bounded functions multiplied by three confining functions in three independent directions of the coordinate space. We notice that if the four μ values take all their integer possible values, then this simple example leads to 4^4 cases.

F. A general degenerate case

Let us assume $F^i_{i\mu\rho} \equiv 0$ and for $i \neq j$ the general degenerate kernels written down in Eq. (6'). As in the previous section, 2.F, even if the solutions are not written in explicit closed form, we can show their confinement properties from their implicit expressions. The proof being of the same type as in 2.F we sketch briefly the elements here whereas the derivation of the solutions is done in Appendix A'.3. The solutions $\hat{K}^i_{j\mu}$ are linear combinations of terms containing $l^i_{j\mu m}$, $g^i_{j\mu m}$, $h^i_{j\mu m}$, and unknown quantities. However these unknown quantities are solutions of algebraic linear systems in such a way that their confinement properties can be studied. Substituting the solutions of these algebraic systems into the unknown quantities present in $\hat{K}^i_{j\mu}$, we can as above show the confinement properties by the same criteria. We finally always have for $\hat{K}^i_{j\mu}$ linear combination of bounded functions multiplied by three functions confining in three different directions of the coordinate space (using the results of Appendix A'.2a, A'.2b, and A'.3).

Let us recall (see the Introduction) that in Eqs. $(1 \pm)$, (2) we could have introduced new variables y_ν's in such a way that the derivatives $\partial/\partial x_\nu$ appear with only three variables, the last becoming a parameter. We have thought it preferable to work with the x_μ's which directly represent either a space or a time coordinate variable.

As in Sec. 2, in order to have in Eq. (2) only three q^i_j's, the kernels of Eq. (11') must be linked.

4. GENERALIZATION OF THE PREVIOUS FORMALISMS[6]

In the Zakharov [5b] formalism (only one coordinate) one of the x_μ is really a variable, the others are parameters. As we have seen previously, using all the x_μ on the same footing leads to a generalization of the I.E. However, for the variable considered Zakharov uses as integration path of the I.E. both $[x_\mu, \infty]$ and $[-\infty, x_\mu]$. Here we extend the previous formalism enlarging the integration path.

(A) First we consider the formalism of Sec. 2. If among the three ψ_j given by Eq. (8), we replace one of them (or two, or three) by

$$\tilde{\psi}_j = \Big(U^0_j(x_1,x_2,x_3,x_4)\delta_{ij} + \frac{1}{1-\rho_j}\int_{-\infty}^{+\infty} K^j_i(y;x_1,\ldots,x_4) \times U^0_j(x_j = y)dy[\theta(y-x_j) + \rho_j\theta(x_j - y)]dy\Big),$$

the ρ_j being real constants and substitute into Eq. $(1 \pm)$, then one can verify that the nine $\{K^i_j\}$ and the six $\{\hat{K}^i_j = K^i_j\ (y = x_j)\}$ still verify the n.l.p.d.e. (10) as well as the same relations (10) linking to the potentials q^j_i. However, now for the K^i_j we have relations for $y \in [-\infty, +\infty]$ instead of only for $y \geqslant x_j$. Secondly we consider instead of Eq. (11) the I.E.

$$K^i_j(y;x_1,x_2,x_3,x_4) = (1-\rho^i_j)\widetilde{F}^i_j(y;x_1,\cdots) + \sum_{m=1}^{3}\int_{-\infty}^{\infty} ds\, F^i_m(s;y;x_1,\cdots)K^m_j(s;x_1,\cdots) \times [\theta(s-x_j) + \rho^i_j\theta(x_j - s)],$$

$$\widetilde{F}^i_j = F^i_j(s = x_j), \quad F^i_j = F^i_j(s - x_j + X_j; y - x_i + X_i; X_4),$$

$$\lim_{|s|\to\infty} F^i_j K^j_m(s,\cdots) = 0,$$

ρ^i_j being constants, F^i_j still satisfying Eq. (13). If we apply $O^{i\pm}_j$ to both sides of this I.E., doing the same algebra as in Appendix A for Eq. (11), it is easy to verify that the $\{K^i_j\}$ still satisfy the n.l.p.d.e. (10). Consequently, from the restrictions $K^i_j(y = x_i)$ we build the q^i_j which are solutions of Eq. (2). Concerning the confinement properties of the solutions, the analysis of Sec. 2 can be extended to the solutions of this I.E.

Secondly, we consider the formalism of Sec. 3. First, if among the ψ_j given by Eq. (8'), we replace one of them (or two, or three) by

$$\tilde{\psi}_j = \Big(U^0_j(x_1,x_2,x_3,x_4)\delta_{ij} + \frac{1}{(1-\rho_{j\mu})}\sum_{\mu\in I_i}\int_{-\infty}^{\infty} dy\, U^0_{j\mu}(y) \times K^j_{i\mu}(y;x_1,\cdots)[\theta(y-x_\mu) + \rho_{j\mu}\theta(x_\mu - y)]\Big),$$

the $\rho_{j\mu}$ being constants, and substitute into (Eq. $1 \pm$), then one can verify that the $\{K^i_{j\mu}\}$ and the $\{\hat{K}^i_{j\mu}\}$ the relations (10'). Here we have relations for $y \in [-\infty, +\infty]$ instead of $y \geqslant x_\mu$ in Sec. 3. Secondly, instead of Eq. (11') we consider the I.E.

$$K^i_j(y;x_1,\ldots,x_4) = [PR{:}1 - [SI{:}\rho{:}i{:}j\mu]][SI{:}F{:}i{:}j\mu](y;x_1,\cdots) + \sum_{m=1}^{3}\sum_{\rho\in I_m}\int_{-\infty}^{+\infty} ds\, F^i_{m\mu\rho}(s;y;x_1,\ldots,x_4) \times K^m_{j\rho}(s;x_1,\cdots)[\theta(s-x_\rho) + \rho^i_{j\mu}\theta(x_\rho - s)],$$

$$\widetilde{F}^i_j = F^i_{j\mu\eta}(s = x_\eta),$$

$$F^i_{j\mu\rho}(\gamma^\rho_{j\mu}(s - x_\rho) + X^\mu_j; y - x_\mu + X^\mu_i; X^\mu_4),$$

$$\lim_{|s|\to\infty} F^i_{j\mu\rho}K^j_{l\rho} = 0,$$

the $\rho^i_{j\mu}$ being constants, $F^i_{j\mu\rho}$ still satisfying Eq. (13'). Applying $O^{i\pm}_{j\mu}$ to both sides of this I.E., doing the same algebra as in Appendix A' for Eq. (11'), we still get that the $\{K^i_{j\mu}\}$ satisfy the n.l.p.d.e. (10'). Consequently, from the restrictions $K^i_{j\mu}(y = x_\mu)$ we build the q^i_j [following the linear combination written down in Eq. (10')] which satisfy Eq. (2). The analysis of Sec. 3 can still be applied to this I.E. in order to show the confinement properties of the reconstructed potentials q^i_j. We emphasize that this I.E. is the more general one of the paper because it contains as subcases all the previous ones.

5. CONCLUSION

In this paper we have taken advantage of our previous

remark[4] that in more than two dimensions, the inversionlike formalism, associated with linear first order differential systems, represents in fact the compatibility condition between different linear differential systems. This paper is an application of this property.

As has been explained in the Introduction, due to the structure of the linear differential parts $l_m = \Sigma a_{m\mu}\partial/\partial x_\mu$ ($m = 1,2,3, \mu = 1,2,3,4$) of the n.l.p.d.e. (2) as well as of the associated linear systems ($1 \pm$), we could have defined a change of variables (mixing the spatial and time coordinate) in such a way that the derivatives are associated with only three new variables whereas the last one appears as a parameter. We have chosen to work with the original form for the l_m because the equation appears in this way in the literature and seems more convenient for discussion of the properties of the solutions directly in the spatial and time coordinates. Nevertheless, there remains, of course, a reflection of this intrinsic feature of Eq. (2) into the formalism presented here. It is in the fact that for the variables X_ν (Sec. 2) or X^μ_ν (Sec. 3), entering into the kernels of the I.E. and in terms of which the solution of Eq. (2) are finally expressed, we have $l_m X_4 = l_m X^\mu_4 = 0$ for all m values whereas $l_m X_j$ and $l_m X^\mu_j$ ($j \leqslant 3$) do not all vanish. However, we notice that these variables X_4, X^μ_4 are necessary in order to really have a three-dimensional coordinate space at fixed time and their associated functions $l^i_j(X_4)$, $l^i_{j\mu}(X^\mu_4)$ play an important role concerning the confinement properties of the solutions at fixed time. The solution of Eq. (2) (corresponding to degenerate kernels of the I.E.) are always obtained as linear combinations of bounded functions multiplied by at least three functions confining in three different directions of the three-dimensional coordinate space. If the $l^i_j(X_4)$, or $l^i_{j\mu}(X^\mu_4)$ were not present, then the solutions will be confined in only two different directions of the space. The richness of the multidimensional formalism is particularly illustrated in Sec. 3 where we show that many different representations of the solutions of the associate linear systems can exist leading to many I.E.'s and consequently to many classes of solutions of Eq. (2). Another striking feature of the solutions of Eq. (2) is the fact that contrary to the one spatial dimensional case, no very particular functionals (like pure exponential) are required to be introduced into the kernels of the I.E. Due to this property, in order to get the confinement property we have, with degenerate kernels of the I.E., only to consider functions of one variable vanishing at infinity. [The same feature was yet present[4] in the two spatial dimensional case of Eq. (2), and we have found[4] a somewhat similar feature for the two spatial dimensional generalization of the nonlinear cubic Schrödinger equation.] Among all these possible solutions of Eq. (2), physical considerations will probably select the interesting ones.

There is a possible class of solutions not considered here (as well as in our previous multidimensional n.l.p.d.e.[4]). They are the rational solutions which could be obtained by limiting process as was explained by Manakov *et al.*[7] in the KdV case starting from pure exponential kernels of the I.E. However, this possibility[8] is neither restricted to these particular kernels nor to the KdV case.

We think that the method presented here for the construction of the solutions of multidimensional n.l.p.d.e. (resulting from the compatibility condition of different linear differential systems) is not restricted to the nonlinear three-wave case and could be extended to other multidimensional n.l.p.d.e.

ACKNOWLEDGMENT

I thank Mark Ablowitz very much for useful conversations.

APPENDIX A

1. n.l.p.d.e. satisfied by the solutions of the I.E. (11)

Let us define $\tilde{F}^i_j = F^i_j(s = x_j, \cdots)$ and consider

$$K^i_j(y;x_1,\ldots,x_y) = \tilde{F}^i_j + \sum_{m=1}^{3} \int_{x_m} F^i_m(s;y;x_1,\ldots,x_4) K^m_j(s;x_1,\ldots,x_y)\,ds, \tag{A1}$$

$$\left(l_p + b_{pi}\frac{\partial}{\partial y} + b_{pj}\frac{\partial}{\partial s}\right) F^i_j(s;y;,\cdots) = 0, \quad p = 1,2,3,$$

$$l_p = \sum_1^4 a_{p\mu}\frac{\partial}{\partial x_\mu}, \tag{A2}$$

with $b_{pj} = a_{pj}$ except $b_{j-1,j} = a_{j-1,j}(\nu_{j-1,j} + 1)$. Recall that $O^{i+}_j = l_j + b_{ji}\partial/\partial y$, $O^{i-}_j = O^{i+}_{j+1}$ and remark from (A2) that $O^{i\pm}_j \tilde{F}^i_j = 0$. We apply $O^{i\pm}_j$ to both sides of (A1):

$$O^{i+}_j K^i_j = -\Sigma a_{j,m}\tilde{F}^i_m \hat{K}^m_j + \sum_m \int_{x_m} (F^i_m l_j K^m_j + K^m_j O^{i+}_j F^i_m)\,ds,$$

$$O^{i-}_j K^i_j = -\Sigma a_{j+1,m}\tilde{F}^i_m \hat{K}^m_j + \sum_m \int_{x_m} (F^i_m l_{j+1} K^m_j + K^m_j O^{i-}_j F^i_m)\,ds.$$

From (A2) we get

$$\left(O^{i+}_j + b_{jm}\frac{\partial}{\partial s}\right) F^i_m = 0,$$

$$\left(O^{i-}_j + b_{j+1m}\frac{\partial}{\partial s}\right) F^i_m = 0$$

and, substituting into the rhs we get

$$O^{i\pm}_j K^i_j - \sum_m \int F^i_m O^{i\pm}_j K^m_j = \sum_m (b_{jm} - a_{jm})\tilde{F}^i_m \hat{K}^m_j$$
$$= a_{j,j+1}\nu_{j,j+1}\tilde{F}^i_{j+1}\hat{K}^{j+1}_j \quad \text{for the } + \text{ case, and}$$
$$= \sum_m (b_{j+1,m} - a_{j+1,m})\tilde{F}^i_m \hat{K}^m_j$$
$$= a_{j+1,j-1}\nu_{j+1,j-1}\tilde{F}^i_{j-1}\hat{K}^{j-1}_j \quad \text{for the } - \text{ case.}$$

If we compare with the solutions of (A1), we get

$$O^{i+}_j K^i_j - a_{j,j+1}\nu_{j,j+1}K^i_{j+1}\hat{K}^{j+1}_j = 0,$$
$$O^{i-}_j K^i_j - a_{j+1,j-1}\nu_{j+1,j-1}K^i_{j-1}\hat{K}^{j-1}_j = 0, \tag{A3}$$

which are the n.l.p.d.e. (10). We recall that $j = 1,2,3$ is a circular permutation ($j + 2 = j - 1$).

2. Dimensional properties of the variables X_j, $j = 1,2,3, X_4$

We recall

$$X_j = \sum_1^3 \beta_{jl} x_l, \quad \sum_{l=1}^3 a_{ml}\beta_{jl} = 0, \quad m = j, j+1,$$
$$\sum_1^3 a_{j-1,l}\beta_{jl} + a_{j-1,j}\nu_{j-1,j} = 0, \tag{A4}$$

$$X_4 = x_4 + \sum_1^3 \beta_{4l} x_l, \quad \sum_1^3 a_{ml}\beta_{4l} + a_{m4} = 0, \quad m = 1,2,3.$$

(i) $\{X_\mu\}, \mu = 1,...,4$ are independent. Otherwise let us assume that there exist $\eta_\mu \not\equiv 0$ such that $\Sigma_1^4 X_\mu \eta_\mu \equiv 0$. Equivalently $\Sigma_{j=1}^3 x_j(\Sigma_1^4 \eta_\mu \beta_{\mu j}) + \eta_4 x_4 \equiv 0$ or $\eta_4 = 0$ and

$$\begin{pmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = 0$$

or (A5)

$$\det(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = 0, \quad \boldsymbol{\beta}_i = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \beta_{i3} \end{pmatrix}, \quad \mathbf{a}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ a_{3i} \end{pmatrix}.$$

However, with (A4) we can calculate directly $\det((\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3))$ and taking into account $\det(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) \neq 0$, we get

$$|\det(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3)| = |\nu_{12}\nu_{23}\nu_{31}a_{12}a_{23}a_{31}[\det(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)]^{-1}|. \tag{A6}$$

(ii) If $t = x_4$, then $X_j, X_i, X_4 - t$ $(i \neq j, i \leqslant 3, j \leqslant 3)$ are independent. Otherwise there exist η_i, η_j, η_4 different of zero such that $\eta_i X_i + \eta_j X_j + \eta_4(X_4 - x_4) \equiv 0$ or

$$\begin{pmatrix} \beta_{ii} & \beta_{ji} & \beta_{4i} \\ \beta_{ij} & \beta_{jj} & \beta_{4j} \\ \beta_{ik} & \beta_{jk} & \beta_{4k} \end{pmatrix} \begin{pmatrix} \eta_i \\ \eta_j \\ \eta_k \end{pmatrix} = 0, \quad i \neq j \neq k \neq i$$

or (A7)

$$\det(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j, \boldsymbol{\beta}_4) = 0.$$

With (A4) we calculate $\det((\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)(\boldsymbol{\beta}_i,\boldsymbol{\beta}_j,\boldsymbol{\beta}_4))$ and get

$$|\det(\boldsymbol{\beta}_i,\boldsymbol{\beta}_j,\boldsymbol{\beta}_4)| = |\nu_{i-1,i}\nu_{j-1,j}a_{i-1,i}a_{j-1,j}a_{k-1,4}[\det(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)]^{-1}|. \tag{A8}$$

(iii) If $t \neq x_4$, for instance $t = x_3$, then $X_i - x_3\beta_{i3}$, $X_j - x_3\beta_{j3}, X_4 - x_3\beta_{43}$ $(i \neq j, i \leqslant 3, j \leqslant 3)$ are independent. The coordinate space is x_1, x_2, x_4; however x_4 appears only in X_4. Consequently, in order to have a three-dimensional coordinate space, we have only to verify that $X_i - x_3\beta_{i3}$, and $X_j - x_3\beta_{j3}$, are independent. Otherwise there exist $\eta_i \neq 0$, $\eta_j \neq 0$ such that $\eta_i(X_i - x_3\beta_{i3}) + \eta_j(X_j - x_3\beta_{j3}) \equiv 0$ or

$$(C_{ij}) = \begin{pmatrix} \beta_{i1} & \beta_{j1} \\ \beta_{i2} & \beta_{j2} \end{pmatrix}, \quad (C_{ij})\begin{pmatrix} \eta_i \\ \eta_j \end{pmatrix} \equiv 0 \quad \text{or} \quad \det(C_{ij}) = 0. \tag{A9}$$

However, with the help of (A4) we can still calculate $\det(C_{ij})$, and we get

$$|\det(C_{ij})| = |\nu_{i-1,i}\nu_{j-1,j}a_{i-1,i}a_{j-1,j}a_{k-1,3}[\det(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)]^{-1}|,$$
$$i \neq j \neq k \neq i, \quad i,j,k \text{ being } 1,2,3, k-1 = k+2. \tag{A10}$$

3. Solutions of (A1) for general degenerate kernels

Let us assume in (A1) $F_i^i = 0$ and

$$F_j^i = \sum_{m=1}^{m_0} g_{jm}^i(s - x_j + X_j)h_{jm}^i(y - x_i + X_i)l_{jm}^i(X_4), \tag{A11}$$

where m_0 is finite. (A1) can be written

$$K_j^i = \widetilde{F}_j^i + \int\int F_j^i F_l^j K_j^i + \int\int F_j^i F_k^j K_j^k + \int F_k^i K_j^k, \quad \text{if } i \leftrightarrow k, K_j^i \leftrightarrow K_j^k. \tag{A12}$$

Let us define

$$\int_0^\infty g_{km}^i(u + X_k)h_{jp}^k(u + X_k) = C_{jk}^{imp}(X_k),$$
$$\int_{x_k}^\infty g_{kp}^i(s - x_k + X_k)K_j^k(s)ds = A_{jkp}^i,$$

and, substituting (A11) into (A12), we get

$$K_j^i(y) = \sum_m l_{jm}^i(X_4)h_{jm}^i(y - x_i + X_i)\Big[g_{jm}^i(X_j) + \sum_p l_{ip}^j(X_4)C_{ij}^{imp}(X_j) \times A_{jip}^j + \sum_p l_{kp}^j(X_4)C_{kj}^{imp}(X_j)A_{jkp}^j\Big]$$
$$+ \sum_m l_{km}^i(X_4)h_{km}^i(y - x_i + X_i)A_{jkm}^i$$

if $i \leftrightarrow k$, then $K_j^i \leftrightarrow K_j^k$. (A13)

The K_j^i defined in (A13) depend upon the set of quantities $\{A_{jip}^j, A_{jkp}^j, A_{jkp}^i, A_{jip}^k\}, p = 1,...,m_0$. These quantities are solutions of algebraic systems which are obtained from (A13) multiplying by g_{ip}^j or g_{ip}^k and integrating. Then two other relations are obtained by $i \leftrightarrow k$,

$$A_{jip}^j\Big[1 - \sum_m l_{ip}^j l_{jm}^i C_{ji}^{jpm} C_{ij}^{imp}\Big] - \sum_{p' \neq p} A_{jip'}^j \sum_m C_{ji}^{jpm} C_{ij}^{imp'} l_{jm}^i l_{ip'}^j$$
$$- \sum_{p'} A_{jkp'}^j \sum_m C_{ji}^{jpm} C_{kj}^{imp'} l_{jm}^i l_{kp'}^j - \sum_{p'} A_{jkp'}^i l_{kp'}^i C_{ki}^{jpp'}$$
$$= \sum_m g_{jm}^i l_{jm}^i C_{ji}^{jpm} \quad \text{and} \quad i \leftrightarrow k, \tag{A14}$$

$$A_{jip}^k - \sum_{p'} A_{jip'}^j \sum_m C_{ji}^{kpm} C_{ij}^{imp'} l_{jm}^i l_{ip'}^i$$
$$- \sum_{p'} A_{jkp'}^j \sum_m C_{ji}^{kpm} C_{kj}^{imp'} l_{kp'}^j l_{jm}^i - \sum_{p'} A_{jkp'}^i l_{kp'}^i C_{ki}^{kpp'}$$
$$= \sum_m g_{jm}^i l_{jm}^i C_{kj}^{kpm} \quad \text{and} \quad i \leftrightarrow k. \tag{A15}$$

Finally $\hat{K}_j^i$ is written

$$\hat{K}_j^i = \sum_m l_{jm}^i(X_4)h_{jm}^i(X_i)\Big[g_{jm}^i(X_j) + \sum_p l_{ip}^i(X_4)C_{ij}^{imp}(X_j)A_{jip}^j + \sum_p l_{kp}^j(X_4)C_{kj}^{imp}(X_j)A_{jkp}^j\Big] + \sum_m l_{km}^i(X_4)h_{km}^i(X_i)A_{jkm}^i.$$

APPENDIX A'

1. n.l.p.d.e. satisfied by the solutions of the I.E. (11')

Let us define $\widetilde{F}_{j\mu}^i = F_{j\mu\rho}^i(s = x_\rho, \cdots)$ and consider

$$K^i_j(y;x_1,\ldots,x_4) = \widetilde{F}^i_{j\mu} + \sum_{m=1}^{3}\sum_{\rho\in I_m}\int_{x_\rho} F^i_{m\mu\rho}(s;y;x_1,\ldots,x_4)$$

$$K^m_{j\rho}(s,x,\ldots,x_4)ds, \qquad (A'1)$$

$$\left(l_p + b^i_{p\mu}\frac{\partial}{\partial y} + b^j_{p\rho}\frac{\partial}{\partial s}\right)F^i_{j\mu\rho}(s;y;x_1,\ldots,x_4) = 0,$$

$$\mu\in I_i,\ \rho\in I_j,\ p = 1,2,3, \qquad (A'2)$$

with $b^i_{p\mu} = a_{p\mu}$ except $b^{p+1}_{p\mu} = a_{p\mu}(\nu_{p\mu}+1)$. We recall that $O^{i+}_{j\mu} = l_j + b^i_{j\mu}\partial/\partial y$, $O^{i-}_{j\mu} = O^{i+}_{j+1,\mu}$ and remark from (A'2) that $O^{i\pm}_{j\mu}\tilde{F}^i_{j\mu} = 0$. We apply $O^{i\pm}_{j\mu}$ to both sides of (A'1)

$$O^{i+}_{j\mu}K^i_{j\mu} = -\sum_m \widetilde{F}^i_{m\mu}\sum_\rho a_{j\rho}\hat{K}^m_{j\rho} + \sum\sum\int (F^i_{m\mu\rho}l_j K^m_{j\rho} + K^m_{j\rho}O^{i+}_{j\mu}F^i_{m\mu\rho})ds,$$

$$O^{i-}_{j\mu}K^i_{j\mu} = -\sum_m \widetilde{F}^i_{m\mu}\sum_\rho a_{j+1\rho}\hat{K}^m_{j\rho} + \sum\sum\int (F^i_{m\mu\rho}l_{j+1}K^m_{j\rho} + K^m_{j\rho}O^{i-}_{j}F^i_{m\mu\rho})ds.$$

From (A'2) we get

$$\left(O^{i+}_{j\mu} + b^m_{j\rho}\frac{\partial}{\partial s}\right)F^i_{m\mu\rho} = 0,$$

$$\left(O^{i-}_{j\mu} + b^m_{j+1\rho}\frac{\partial}{\partial s}\right)F^i_{m\mu\rho} = 0,$$

and, substituting into the rhs, we get

$$O^{i\pm}_{j\mu}K^i_{j\mu} - \sum_m\sum_{\rho\in I_m}\int F^i_{m\mu\rho}O^m_{j\rho}K^m_{j\rho}$$

$$= \sum_m\sum_{\rho\in I_m}\widetilde{F}^i_{m\mu}\hat{K}^m_{j\rho}(-a_{j\rho}+b^m_{j\rho})$$

$$= \widetilde{F}^i_{j+1\mu}\sum_{\rho\in I_{j+1}} a_{j\rho}\nu_{j\rho}\hat{K}^{j+1}_{j\rho} \quad \text{for the } + \text{ case and}$$

$$= \sum_m\sum_{\rho\in I_m}\widetilde{F}^i_{m\mu}(-a_{j+1\rho}+b^m_{j+1\rho})\hat{K}^m_{j\rho}$$

$$= \widetilde{F}^i_{j-1,\mu}\sum_{\rho\in I_j} a_{j+1,\rho}\nu_{j+1,\rho}\hat{K}^{j-1}_{j\rho} \quad \text{for the } - \text{ case.}$$

Comparing with the solutions of (A'1) we get Eq. (10') [$j = 1,2,3$ is a circular permutation ($j+1 = j-2$)].

2.a. Dimensional properties of the variables X^μ_j, $j = 1,2,3$, X^μ_4, μ fixed, written down in Eq. (4b")

(i) $\{X^\mu_\alpha\}$, $\alpha = 1,\ldots,4$, are independent. Otherwise let us assume that there exist $\eta_\alpha \not\equiv 0$ such that $\Sigma_1^4 X^\mu_\alpha\eta_\alpha = 0$. Equivalently, $\eta_4 = 0$ and

$$\begin{pmatrix}\beta^\mu_{1\varphi_1} & \beta^\mu_{2\varphi_1} & \beta^\mu_{3\varphi_1}\\ \beta^\mu_{1\varphi_2} & \beta^\mu_{2\varphi_2} & \beta^\mu_{3\varphi_2}\\ \beta^\mu_{1\varphi_3} & \beta^\mu_{2\varphi_3} & \beta^\mu_{3\varphi_3}\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\end{pmatrix} = 0$$

or

$$\left.\begin{aligned}&\det(\beta^\mu_1\beta^\mu_2\beta^\mu_3) = 0\\ &\varphi_i\neq\mu, \quad i = 1,2,3\end{aligned}\right\}, \quad \beta^\mu_i = \begin{pmatrix}\beta^\mu_{i\varphi_1}\\ \beta^\mu_{i\varphi_2}\\ \beta^\mu_{i\varphi_3}\end{pmatrix}.$$

We define

$$\mathbf{a}_{\varphi_i} = \begin{pmatrix}a_{1\varphi_i}\\ a_{2\varphi_i}\\ a_{3\varphi_i}\end{pmatrix},$$

with (4b") we get

$$|\det(\beta^\mu_1\beta^\mu_2\beta^\mu_3)| = |a_{1\mu}a_{2\mu}a_{3\mu}\nu_{1\mu}\nu_{2\mu}\nu_{3\mu}|\,[\det(\mathbf{a}_{\varphi_1}\mathbf{a}_{\varphi_2}\mathbf{a}_{\varphi_3})]^{-1}.$$

(ii) If $t = x_\mu$, then X^μ_j, X^μ_i, $X^\mu_4 - t$ ($i\neq j$, $i\leqslant 3, j\leqslant 3$) are independent. Otherwise there exist η_i,η_j,η_4 such that $\eta_i X^\mu_i + \eta_j X^\mu_j + \eta_4(X^\mu_4 - x_\mu) = 0$ or

$$\begin{pmatrix}\beta^\mu_{i\varphi_1} & \beta^\mu_{j\varphi_1} & \beta^\mu_{4\varphi_1}\\ \beta^\mu_{i\varphi_2} & \beta^\mu_{j\varphi_2} & \beta^\mu_{4\varphi_2}\\ \beta^\mu_{i\varphi_3} & \beta^\mu_{j\varphi_3} & \beta^\mu_{4\varphi_3}\end{pmatrix}\begin{pmatrix}\eta_i\\ \eta_j\\ \eta_4\end{pmatrix} = 0$$

or

$$\det(\beta^\mu_i\beta^\mu_j\beta^\mu_4) = 0, \quad \varphi_i\neq\mu, \quad i = 1,2,3.$$

With (4b") we get

$$|\det(\beta^\mu_i\beta^\mu_j\beta^\mu_4)| = |a_{1\mu}a_{2\mu}a_{3\mu}\nu_{i-1,\mu}\nu_{j-1,\mu}|\,[\det(\mathbf{a}_{\varphi_1}\mathbf{a}_{\varphi_2}\mathbf{a}_{\varphi_3})]^{-1}.$$

(iii) If $t\neq x_\mu$, for instance, $t = x_{\varphi_3}$, then $X^\mu_i - x_{\varphi_3}\beta^\mu_{i\varphi_3}$, $X^\mu_j - x_{\varphi_3}\beta^\mu_{j\varphi_3}$, $X_4 - x_{\varphi_3}\beta^\mu_{4\varphi_3}$ ($i\neq j$, $i\leqslant 3, j\leqslant 3$, $\varphi_i\neq\mu$) are independent. x_μ appears only in $X_4 - x_{\varphi_3}\beta^\mu_{4\varphi_3}$ and we have to verify that $\eta_i\neq 0$, $\eta_j\neq 0$ and $\eta_i(X^\mu_i - x_{\varphi_3}\beta^\mu_{i\varphi_3}) + \eta_j(X^\mu_j - x_{\varphi_3}\beta^\mu_{j\varphi_3}) = 0$ is not possible; otherwise,

$$(C^\mu_{ij}) = \begin{pmatrix}\beta^\mu_{i\varphi_1} & \beta^\mu_{j\varphi_1}\\ \beta^\mu_{i\varphi_2} & \beta^\mu_{j\varphi_2}\end{pmatrix}, \quad (C^\mu_{ij})\begin{pmatrix}\eta_i\\ \eta_j\end{pmatrix} = 0$$

or

$$\det(C_{ij}) = 0.$$

However, with (4b") we get ($i\neq j\neq k\neq i$, $k\leqslant 3$)

$$|\det(C^\mu_{ij})| = |\nu_{i-1,\mu}\nu_{j-1,\mu}a_{i-1,\mu}a_{j-1,\mu}a_{k-1,\varphi_3} \times [\det(\mathbf{a}_{\varphi_1}\mathbf{a}_{\varphi_2}\mathbf{a}_{\varphi_3})]^{-1}|.$$

2.b. Dimensional properties of the variables $X^{\mu_1}_i$, $X^{\mu_2}_j, X^{\mu_1}_4$, at fixed time t and $\mu_1\neq\mu_2$, i≠j

We recall

$$X^{\mu_1}_i = \sum_{\varphi_i\neq\mu_1}\beta^{\mu_1}_{i\varphi_i}x_{\varphi_i}, X^{\mu_1}_4 = x_{\mu_1} + \sum_{\varphi_i\neq\mu_1}\beta^{\mu_1}_{4\varphi_i}x_{\varphi_i},$$

$$X^{\mu_2}_j = \beta^{\mu_2}_{j\varphi_2}x_{\varphi_2} + \beta^{\mu_2}_{j\mu_1}x_{\varphi_1} + \beta^{\mu_2}_{j\mu_1}x_{\varphi_1}, \quad i = 1,2,3, j = 1,2,3,$$

$\varphi_i\neq\mu_1$, $\varphi_i = \varphi_1,\varphi_2,\varphi_3,\mu_2 = \varphi_1$.

We define

$$\beta_{\mathrm{I}} = \begin{pmatrix}\beta^{\mu_1}_{i\mu_2} & 0 & \beta^{\mu_1}_{4\mu_2}\\ \beta^{\mu_1}_{i\varphi_2} & \beta^{\mu_2}_{j\varphi_2} & \beta^{\mu_1}_{4\varphi_2}\\ \beta^{\mu_1}_{i\varphi_3} & \beta^{\mu_2}_{j\varphi_3} & \beta^{\mu_1}_{4\varphi_3}\end{pmatrix},$$

$$\beta_{\mathrm{II}} = \begin{pmatrix}\beta^{\mu_1}_{i\varphi_2} & \beta^{\mu_2}_{j\varphi_2} & \beta^{\mu_2}_{4\varphi_2}\\ \beta^{\mu_1}_{i\varphi_3} & \beta^{\mu_2}_{j\varphi_3} & \beta^{\mu_1}_{4\varphi_3}\\ 0 & \beta^{\mu_2}_{j\mu_1} & 1\end{pmatrix},$$

$$\beta_{\rm III}=\begin{pmatrix}\beta^{\mu_1}_{i\mu_2} & 0 & \beta^{\mu_1}_{4\varphi_2}\\ \beta^{\mu_1}_{i\varphi_3} & \beta^{\mu_2}_{j\varphi_3} & \beta^{\mu_1}_{4\varphi_3}\\ 0 & \beta^{\mu_2}_{j\mu_1} & 1\end{pmatrix}.$$

(i) If $t=x_{\mu 1}$, the remaining spatial coordinates define a three-dimensional space. Otherwise there exist $\eta_i\neq 0$, $\eta_j\neq 0$, $\eta_4\neq 0$ such that

$$\eta_i X_i^{\mu_1}+\eta_j(X_j^{\mu_2}-\beta^{\mu_2}_{j\mu_1}x_{\mu_1})+\eta_4(X_4^{\mu_1}-x_{\mu_1})=0$$

or

$$\det\beta_{\rm I}=0.$$

However, with (4b") we get, $i\neq j\neq k\neq i$, $k\leqslant 3$,

$$|\det\beta_{\rm I}|=|a_{i-1,\mu_1}a_{j-1,\mu_2}a_{k-1,\mu_1}\nu_{i-1,\mu_1}\nu_{j-1,\mu_2}\times[\det(\mathbf{a}_{\varphi_1}\mathbf{a}_{\varphi_2}\mathbf{a}_{\varphi_1})]^{-1}|.$$

(ii) If $t=x_{\mu_2}$, we still have a three-dimensional space. Otherwise there exist $\eta_i\neq 0$, $\eta_j\neq 0$, $\eta_4\neq 0$ such that

$$\eta_i(X_i^{\mu_1}-\beta^{\mu_1}_{i\mu_2}x_{\mu_2})+\eta_j X_j^{\mu_2}+\eta_4(X_4^{\mu_1}-\beta^{\mu_1}_{4\mu_2}x_{\mu_2})=0$$

or

$$\det\beta_{\rm II}=0.$$

A direct calculation gives, $i\neq j\neq k\neq i$, $k\leqslant 3$,

$$|\det\beta_{\rm II}|=|a_{i-1,\mu_1}a_{j-1,\mu_2}a_{k-1,\mu_2}\nu_{i-1,\mu_1}\nu_{j-1,\mu_2}\times[\det(\mathbf{a}_{\varphi_1}\mathbf{a}_{\varphi_2}\mathbf{a}_{\varphi_1})]^{-1}|.$$

(ii) If t is neither x_{μ_1} or x_{μ_2} but for instance x_{φ_2}, we still have a three-dimensional space. Otherwise there exist $\eta_i\neq 0$, $\eta_j\neq 0$, $\eta_4\neq 0$ with $\eta_i(X_i^{\mu_1}-\beta^{\mu_1}_{i\varphi_2}x_{\varphi_2})+\eta_j(X_j^{\mu_2}-\beta^{\mu_2}_{j\varphi_2}x_{\varphi_2})+\eta_4(X_4^{\mu_1}-\beta^{\mu_1}_{4\varphi_2}x_{\varphi_2})=0$

or

$$\det\beta_{\rm III}=0.$$ With (4b") we get

$$|\det\beta_{\rm III}|=|a_{i-1,\mu_1}a_{j-1,\mu_2}a_{k-1,\varphi_2}\nu_{i-1,\mu_1}\nu_{j-1,\mu_2}[\det(\mathbf{a}_{\varphi_1}\mathbf{a}_{\varphi_2}\mathbf{a}_{\varphi_1})]^{-1}|.$$

3. Solutions of (A'1) for general degenerate kernels

Among the elements $\{1,2,3,4\}$ let us call ρ_i those belonging to I_i, or $\rho_i\in I_i$ and similarly $\rho_j\in I_j$, $\rho_k\in I_k$. We assume in (A'1) $F^i_{i\mu\rho}\equiv 0$ and

$$F^i_{j\rho_i\rho_j}=\sum_{m=1}^{m_0}g^{i\rho_j}_{j\rho_i m}(\gamma^{\rho_i}_{j\rho_i}(s-x_{\rho_j})+X_j^{\rho_i})\,h^i_{jpj,m}(y-x_{\rho_i}+X_i^{\rho_i})\,l^i_{j\rho_i m}(X_4^{\rho_i}),\qquad (A'3)$$

where m_0 is finite and $g^{i\rho_j}_{j\rho_i m}(s=x_{\rho_j})=g^i_{j\rho_i m}(X_j^{\rho_i})$. (A'1) can be written

$$K^i_{j\rho_i}=\widetilde{F}^i_{j\rho_i}+\sum_{\rho_j\in I_j}\int\int F^i_{j\rho_i\rho_j}\Big\{\sum_{\rho_i'\in I_i}F^j_{i\rho_j\rho_i'}K^i_{j\rho_i'}+i\leftrightarrow k\Big\}+\sum_{\rho_k\in I_k}\int F^i_{k\rho_i\rho_k}K^k_{j\rho_k}\quad\text{if } i\leftrightarrow k,\ K^i_{j\rho_i}\leftrightarrow K^k_{j\rho_k}.$$

Let us define

$$C^{imp}_{jk\rho_i\rho_k}(X^{\rho_i}_k,X^{\rho_k}_k)=\int_0^\infty g^{i\rho_k}_{k\rho_i m}(\gamma^{\rho_k}_{k\rho_i}u+X^{\rho_i}_k)\,h^k_{j\rho_k p}(u+X^{\rho_k}_k)\,du,$$

$$A^{i\rho_i}_{jkm}=\sum_{\rho_k\in I_k}\int_{x_{\rho_k}}^\infty g^{i\rho_k}_{k\rho_i m}(\gamma^{\rho_k}_{k\rho_i}(s-x_{\rho_k})+X^{\rho_i}_k)K^k_{j\rho_k}(s)\,ds$$

and substituting (A'3) into (A'4), we get two equations, where $K^i_{j\rho_i},K^k_{j\rho_k}$ depend upon the set of quantities $\{A^{j\rho_j}_{jip},A^{k\rho_k}_{jip},A^{j\rho_j}_{jkp},A^{i\rho_i}_{jkp}\}$, $p=1,\ldots,m_0$, $\rho_i\in I_i$, $\rho_j\in I_j$, $\rho_k\in I_k$. These quantities are solutions of algebraic systems which are obtained from (A'4) multiplying by $g^{j\rho_i}_{i\rho_i p}$ or $g^{k\rho_i}_{i\rho_i p}$, integrating, and summing over all the $\rho_i\in I_i$ (and $i\leftrightarrow k$):

$$A^{j\rho_j}_{jip}-\sum_m\sum_{\rho_i\in I_i}l^i_{j\rho_i m}C^{jpm}_{ji\rho_j\rho_i}\sum_{m'}\sum_{\rho_j'\in I_j}\{l^i_{i\rho_j' m'}C^{imm'}_{ij\rho_i\rho_j'}A^{j\rho_j'}_{jim'}+l^j_{k\rho_j' m'}C^{imm'}_{kj\rho_i\rho_j'}A^{j\rho_j'}_{jkm'}\}-\sum_m\sum_{\rho_i\in I_i}l^i_{k\rho_i m}C^{jpm}_{ki\rho_j\rho_i}A^{i\rho_i}_{jkm}=\sum_m\sum_{\rho_i\in I_i}l^i_{j\rho_i m}g^i_{j\rho_i m}C^{jpm}_{ji\rho_j\rho_i}\quad\text{and}\quad i\leftrightarrow k,\qquad (A'5)$$

$$A^{k\rho_k}_{jip}-\sum_m\sum_{\rho_i\in I_i}l^i_{j\rho_i m}C^{kpm}_{ji\rho_k\rho_i}\sum_{m'}\sum_{\rho_j\in I_j}[l^i_{k\rho_i m'}C^{imm'}_{ij\rho_i\rho_j}A^{j\rho_j}_{jim'}+l^j_{k\rho_j m'}A^{j\rho_j}_{jkm'}C^{imm'}_{kj\rho_i\rho_j}]-\sum_m\sum_{\rho_i\in I_i}l^i_{k\rho_i m}C^{kpm}_{ki\rho_k\rho_i}A^{i\rho_i}_{jkm}=\sum_m\sum_{\rho_i\in I_i}l^i_{j\rho_i m}g^i_{j\rho_i m}C^{kpm}_{ji\rho_i\rho_i}\quad\text{and}\quad i\leftrightarrow k.\qquad (A'6)$$

Finally $\hat{K}^i_{j\rho_i}$ is written

$$\hat{K}^i_{j\rho_i}=\sum_m l^i_{j\rho_i m}(X_4^{\rho_i})\,h^i_{j\rho_i m}(X_i^{\rho_i})\Big\{g^i_{j\rho_i m}(X_j^{\rho_i})+\sum_{m'}\sum_{\rho_i\in I_j}l^j_{i\rho_j m'}\times(X_4^{\rho_j})C^{imm'}_{ij\rho_i\rho_j}A^{j\rho_j}_{jim'}+l^j_{k\rho_j m'}(X_4^{\rho_j})C^{imm'}_{kj\rho_i\rho_j}A^{j\rho_j}_{jkm'}\Big\}+\sum_m l^i_{k\rho_i m}(X_4^{\rho_i})\,h^i_{k\rho_i m}(X_i^{\rho_i})A^{i\rho_i}_{jkm}.\qquad (A'7)$$

Now we assume that the kernels in (A'3) are of the Eq. (6') type and always disregard the zeros of the determinant of (A'5), (A'6). From (A'5), (A'6) we see that the solutions are linear combination of bounded functions multiplied by $l^i_{j\rho_i m}g^i_{j\rho_i m}$ and $i\leftrightarrow k$. Substituting into $\hat{K}^i_{j\rho_i}$, we still get a linear combination of bounded terms multiplied by functions confining either in $X_4^{\mu_1},X_j^{\mu_1},X_i^{\mu_2}$ or in $X_4^{\mu_1},X_j^{\mu_2},X_i^{\mu_1}$ with $i\neq j$.

[1] P.D. Lax, Com. Pure Appl. Math. **21**, 457 (1968); M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1962 (1973); M.J. Ablowitz and R. Haberman, J. Math. Phys. **16**, 2301 (1975); for the Yang–Mills fields see A.A. Belavin and V.E. Zakharov, *Mathematical Problems in Theoretical Physics* (Springer-Verlag, Berlin, 1978), Vol. 80, p. 229, and many lectures: For instance, A. Degasperis and F. Calogero: "Lectures at the International Advanced Study Institute on Non Linear Equations in Physics and Mathematics," Istanbul, August 1977. See also D.V. Choodnosky and G.V. Choodnosky "Séminaire sur les équations non linéaires," Ecole Polytechnique, 1977–1978.

[2]V.E. Zakharov and A.B. Shabat, Funct. Anal. Appl. **8**, 226 (1974); M.J. Ablowitz and R. Haberman, Phys. Rev. Lett. **18**, 1185 (1975).
[3]H. Cornille, "Inversion-like Integral Equations" (Problems Inverses, Montpellier, 1977), and J. Math. Phys. **19**, 1463 (1978).
[4]H. Cornille, J. Phys. A **11**, 1509 (1978), Saclay preprint DPh-T 78/56, and J. Math. Phys. **20**, 199 (1978).
[5a]In the one-dimensional case see, for instance, V.E. Zakharov and S.V. Manakov, Zh. Eksp. Teor. Fiz. **69**, 1654 (1975) [Sov. Phys. JETP **42**, 842 (1975)]; D.J. Kaup, Stud. Appl. Math. **55**, 9 (1976); H. Cornille, J. Math. Phys. **19**, 1978.
[5b]V.E. Zakharov, Sov. Phys. Dokl. **21**, 322 (1976).
[6]M. Ablowitz has suggested to me the extensions of the Secs. 2 and 3 for both the representations (8), (8′) of the solutions of the linear systems $(1 \pm)$ and for the I.E. (11), (11′) in the cases where the integrations paths are $[-\infty, x_\mu]$ and $[x_\mu, \infty]$.
[7]S.V. Manakov *et al.*, Phys. Lett. A **63**, 205 (1977); see also M.J. Ablowitz and J. Satsuma, J. Math. Phys. **19**, 2180 (1978).
[8]H. Cornille, to appear.

Rigid body motions, space curves, prolongation structures, fiber bundles, and solitons

M. Lakshmanan

Department of Physics, University of Madras, Post-Graduate Center, Tiruchirapalli—620 020, Tamilnadu, India
(Received 5 April 1978)

The dynamics of a nonlinear string of constant length represented by a helical space curve may be studied through a consideration of the motion of an arbitrary rigid body along it. The resulting set of compatibility equations is shown to result in the class of nonlinear evolution equations solvable through the two component inverse scattering phenomenology. A class of pseudopotentials and prolongation structures follow naturally due to the intrinsic group structure of the phenomenon. This leads to an identification of the underlying fiber bundle structure and connection forms. Thus a unified picture emerges for a class of soliton possessing evolution equations.

1. INTRODUCTION

Recent advances in the solution procedure of nonlinear partial differential equations have made it possible to solve the initial value problem of a large class of evolution equations.[1-4] These are especially of dispersive type in one-space one-time dimensions and exactly solvable through the inverse scattering transform[1,2] and Bäcklund transformation methods.[3,4] Broadly they may be identified[2] with the well-known AKNS system (and its generalizations) for the two dependent variables $q(x,t)$ and $r(x,t)$,

$$A_x = qC - rB, \tag{1a}$$

$$q_t - 2Aq = B_x + 2i\zeta B, \tag{1b}$$

$$r_t + 2Ar = C_x - 2i\zeta C. \tag{1c}$$

Equation (1) is the result of imposition of the integrability condition on the two-component eigenvalue problem

$$v_{1x} + i\zeta v_1 = qv_2, \tag{2a}$$

$$v_{2x} - i\zeta v_2 = rv_1, \tag{2b}$$

and the time evolution of the eigenfunctions:

$$v_{1t} = A(x,t,\zeta)v_1 + B(x,t,\zeta)v_2, \tag{3a}$$

$$v_{2t} = C(x,t,\zeta)v_1 - A(x,t,\zeta)v_2, \tag{3b}$$

with the usual identifications.[2]

Now as all the solvable evolution equations associated with (1) possess analogous solitonic and other integrability properties, a basic question arises: What is the underlying mechanism that aids in their solvability and what possible interpretation of the equations and their solutions can be given. A number of apparently differing viewpoints have been offered during the past few years.[5-12] From an elementary geometric point of view, Lamb[7] has demonstrated that the intrinsic equations governing the motion of helical space curves are intimately related to the equations of type (1). Stressing an abstract geometric point of view, Wahlquist and Estabrook[5] have shown that the existence of certain Lie-algebraic structures—prolongation structures—and pseudopotentials are the basic mechanisms. These structures have in turn further been interpreted by Hermann[6] as connection forms of certain fiber bundles. He had also shown that the Lax's equation of nonlinear theory is intimately related to the rigid body equations and could be interpreted as Poisson–Moyal brackets.[10] Several of these concepts have further been elucidated by Morris,[8] Corones,[9] and Dodd and Gibbon[11] and others. Other related geometric or group-theoretical pictures of some of these equations also exist.[12,13]

It is remarkable that in all these approaches either geometry or group structure or both seem to play a predominant role. This makes one suspect that these various ideas could possibly be related to one another through a single unified phenomenon. In this paper we wish to investigate this interrelationship and show that these different views could be seen as the many facets of a single picture of rigid body motion along nonlinear strings or space curves. Consider the dynamics of a homogeneous flexible nonlinear string of constant length attached to some elastically yielding support.[14] It could be studied starting from a suitable Lagrangian as done by Broer.[14] We can also consider the string to be represented by a helical space curve. At an arbitrary point along the string an arbitrary rigid body could be attached. Then the motion of the string and the body are compatible, a consideration which is shown to result in the solvable evolution Eq. (1) (Sec. 3). Further the Serret–Frenet equations describing the space curves and the equations of motion of the rigid body have an SO(3) structure. Both these sets could be reduced to two Riccati equations, in terms of a Darboux vector, having an SL(2,R) structure. A further transformation of these equations results in the eigenvalue problem of the type (2) and (3) having an SU(2) structure. Translating these results in the language of differential forms of exterior algebra, one arrives at the notions of pseudopotentials and prolongation structures in terms of simple geometrical and physical quantities (Sec. 4). The associated fiber bundle structure and connections could also be worked out without difficulty. The space of variables defining the string then forms the base space of the bundle, while the fiber is defined as a space of variables of the rigid body. Thus a unified picture emerges. This is satisfactory, as the study of strings and rigid bodies, which lie at the heart of classical physics, leads to the most recent developments of theoretical physics.

To be self-contained we present the salient features of the theory of space curves and rigid body dynamics in the next section.

 0022-2488/79/081667-06$01.00

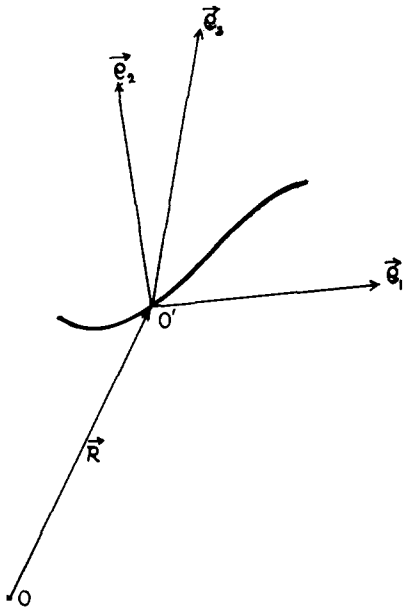

FIG. 1a. A helical space curve in E^3.

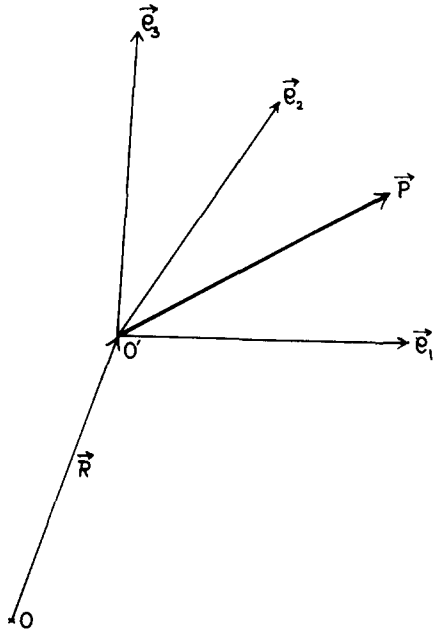

FIG. 1b. A rigid body.

2. SPACE CURVES AND RIGID BODY DYNAMICS

A. Theory of space curves.[15]

A *regular space* curve in E^3 is defined as the locus of the end points of a vector $\mathbf{R}(x)\equiv(r_1(x),r_2(x),r_3(x))$ $(x_0<x<x_1)$ (see Fig. 1a) and that $\mathbf{R}(x)$ obeys certain regularity conditions. The *tangent vector* is then identified with $\mathbf{R}'\equiv d\mathbf{R}/dx = \mathbf{R}_x$ at the point 0'. From the definition of *arc length* $s(x)$ of a curve from a fixed point x_0 to a variable point x we have

$$s(x) = \int_{x_0}^{x} [\mathbf{R}'(x)\cdot\mathbf{R}'(x)]^{1/2}\,dx, \tag{4}$$

and so

$$\frac{ds(x)}{dx} = |\mathbf{R}'(x)|. \tag{5}$$

Then it is clear that if we parametrize the curve with the arc length itself (which we hereafter identify with x itself) the tangent vector $\mathbf{R}'$ becomes a unit vector,

$$\mathbf{R}'(x) = \mathbf{e}_1(x). \tag{6}$$

Now defining the total curvature as the rate of change of the direction of the tangent vector with respect to a fixed direction as $\mathbf{R}(x)$ moves along the arc,

$$\kappa_T = \int_{x_0}^{x_1} [\mathbf{R}''\cdot\mathbf{R}'']^{1/2}\,dx, \tag{7}$$

the concept of *curvature* at each point x is introduced as

$$\kappa(x) = |\mathbf{R}''| = |\mathbf{e}_1'|. \tag{8}$$

As $\mathbf{e}_1'\cdot\mathbf{e}_1 = 0$, a second unit normal called the *principal normal* is chosen such that

$$\mathbf{e}_1' = \kappa\mathbf{e}_2 \quad (\kappa>0). \tag{9}$$

Then the right-handed unit orthogonal trihedran can be completed by defining the third unit vector called the *binormal* by

$$\mathbf{e}_3 = \mathbf{e}_1\times\mathbf{e}_2. \tag{10}$$

Our description could now be completed by introducing the concept of *torsion*. From the definition of total torsion, which is the rate of change of direction of $\mathbf{e}_3$ with respect to a fixed direction as $\mathbf{R}$ moves,

$$\tau_T = \int_{x_0}^{x_1} [\mathbf{e}_3'\cdot\mathbf{e}_3']^{1/2}\,dx, \tag{11}$$

the torsion at any point x is given by

$$\tau(x) = |\mathbf{e}_3'|. \tag{12}$$

From the relation $\mathbf{e}_3'\cdot\mathbf{e}_3 = 0$, $\mathbf{e}_3'\cdot\mathbf{e}_1 = 0$ it then follows that

$$\mathbf{e}_3' = \pm\tau\mathbf{e}_2. \tag{13}$$

We choose the $-$ sign for convenience. Since any other vector can be expressed in terms of $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, we have

$$\mathbf{e}_2' = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3 \tag{14}$$

and from the orthonormality relations we find $a = -\kappa$, $b = 0$, $c = \tau$. Thus we have finally the system of three equations known as the Serret–Frenet equations,

$$\begin{bmatrix}\mathbf{e}_1'\\ \mathbf{e}_2'\\ \mathbf{e}_3'\end{bmatrix} = \begin{bmatrix}0 & \kappa & 0\\ -\kappa & 0 & \tau\\ 0 & -\tau & 0\end{bmatrix}\cdot\begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \mathbf{e}_3\end{bmatrix}. \tag{15a}$$

or briefly, in matrix form

$$\underline{\mathbf{e}}' = \mathsf{F}(\kappa,\tau)\underline{\mathbf{e}},\quad \underline{\mathbf{e}}^T = (\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3), \tag{15b}$$

which forms the basis of the theory of space curves. The space curve is then defined uniquely within congruence, i.e., within rigid motions.

B. Rigid body dynamics

It is well known that a rigid body can be thought of as consisting of a moving orthogonal trihedral of unit vectors fixed to it (Fig. 1b). A complete description of the motion of the body can then be obtained from a discussion of the mo-

tion of the trihedral fixed to it. When the trihedral moves, let the origin 0′ of it be displaced in time t (with respect to a fixed coordinate system of origin 0) as per

$$\frac{d\mathbf{R}}{dt} \equiv \mathbf{R}_t = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2 + \gamma \mathbf{e}_3. \tag{16}$$

Here α, β, and γ are specified functions. If now $\mathbf{P}$ is assumed to be a vector from 0′ to any point fixed in the body, then the position $\mathbf{Q}$ of this point with respect to the fixed system of coordinates at origin 0 is given by

$$\mathbf{Q} = \mathbf{R} + \mathbf{P}. \tag{17}$$

Then the velocity of the point $\mathbf{Q}$ is

$$\mathbf{Q}_t = \mathbf{R}_t + \mathbf{P}_t, \tag{18}$$

where $\mathbf{R}_t$ is as given in Eq. (16). Thus the velocity of $\mathbf{Q}$ is the vector sum of the velocity of the origin 0′ with respect to 0 and the velocity of $\mathbf{Q}$ if 0 were to coincide with 0′ and the rigid body simply rotated about it with an angular velocity $\boldsymbol{\omega}$ independent of the position of $\mathbf{P}$;

$$\mathbf{P}_t = \boldsymbol{\omega} \times \mathbf{P}, \tag{19}$$

where

$$\boldsymbol{\omega} = \omega_1 \mathbf{e}_1 + \omega_2 \mathbf{e}_2 + \omega_3 \mathbf{e}_3. \tag{20}$$

Since the vector $\mathbf{P}$ is quite arbitrary, this includes also $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. Thus we have for the time variation of the trihedral

$$\mathbf{e}_{it} = \boldsymbol{\omega} \times \mathbf{e}_i \quad (i = 1,2,3), \tag{21}$$

or

$$\begin{bmatrix} \mathbf{e}_{1t} \\ \mathbf{e}_{2t} \\ \mathbf{e}_{3t} \end{bmatrix} = \begin{bmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{bmatrix}. \tag{22}$$

In matrix form we write this as

$$\underline{\mathbf{e}}_t = \mathsf{G}(\omega_j)\underline{\mathbf{e}}, \quad \underline{\mathbf{e}}^T = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3). \tag{23}$$

Further, we could introduce the quantities

$$\mathbf{e}_i = \sum_{j=1}^{3} \mathbf{e}_{ij} \mathsf{X}_j \quad (i,j = 1,2,3), \tag{24}$$

where $\mathsf{X}_1, \mathsf{X}_2, \mathsf{X}_3$ are the (3×3) matrix representations of the infinitesimal generators of the rotation group SO(3) obeying the commutation relations

$$[\mathsf{X}_i, \mathsf{X}_j] = -\epsilon_{ijk} \mathsf{X}_k \quad (i,j,k = 1,2,3). \tag{25}$$

Then Eq. (21) could be rewritten as

$$\mathbf{e}_{it} = [\mathbf{e}_i, \mathsf{G}(\omega_j)], \tag{26}$$

where the $\mathsf{G}(\omega_j)$ of Eq. (23) could be expressed as

$$\mathsf{G}(\omega_j) = \omega_1 \mathsf{X}_1 + \omega_2 \mathsf{X}_2 + \omega_3 \mathsf{X}_3. \tag{27}$$

3. MOTION OF RIGID BODY ALONG A HELICAL SPACE CURVE AND THE NONLINEAR EVOLUTION EQUATIONS

Now we assume that the helical space curve corresponds to say, a nonlinear string of constant length attached to flexible supports. The arbitrary trihedral at the point $\mathbf{R}(x)$ could now be replaced by an arbitrary rigid body, in lieu of the properties quoted in Sec. 2 B. When the body is set in motion, the string will also move as the body is constrained to the string. Accordingly we assume hereafter $\mathbf{R} = \mathbf{R}(x,t)$ and $\mathbf{e}_i(x,t)$ with x, the arc length denoting the position of the body along the string. Similarly the other quantities specifying the system are also functions of both x and t, i.e., $\omega_i(x,t)$, $\alpha(x,t)$, $\beta(x,t)$, and $\gamma(x,t)$.

The compatibility of the motion of the string and the body may be expressed with reference to the points $\mathbf{Q}$ of the body (vide Sec. 2 B),

$$\mathbf{Q}_{xt} = \mathbf{Q}_{tx}. \tag{28}$$

Since $\mathbf{Q} = \mathbf{P} + \mathbf{R}$, we have two separate conditions

$$\mathbf{P}_{xt} = \mathbf{P}_{tx} \rightarrow \mathbf{e}_{ixt} = \mathbf{e}_{itx} \quad (i = 1,2,3) \tag{29}$$

and

$$\mathbf{R}_{xt} = \mathbf{R}_{tx}. \tag{30}$$

Substituting now the Serret–Frenet equations (15) and the rigid body equations (22) in the above, the following six equations for the eight unknowns $\kappa, \tau, \omega_1, \omega_2, \omega_3, \alpha, \beta$, and γ result:

$$\mathsf{F}_t - \mathsf{G}_x + \mathsf{FG} - \mathsf{GF} = 0, \tag{31a}$$

or

$$\kappa_t = \omega_{3x} - \tau\omega_2, \quad \tau_t = \omega_{1x} + \kappa\omega_1, \quad \omega_{2x} = \kappa\omega_1 - \tau\omega_2, \tag{31b}$$

and

$$\begin{bmatrix} \alpha_x \\ \beta_x \\ \gamma_x \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 \\ -\omega_3 \\ -\omega_2 \end{bmatrix}. \tag{32}$$

Solving for Eqs. (31) and substituting them in (32), α, β, and γ may be obtained. The above six equations (31) and (32) can be shown to be equivalent to the ones that were obtained by Lamb[7] for the motion of helical curves and by Corones for the integrability of pseudopotentials.[9b] These set of equations are also essentially equivalent to the ones that were obtained by Broer[14] describing the dynamics of a string of constant length through a variational principle involving stress and strain of the string.

The system of equations (31) and (32) forms an incomplete set. So one can choose the ω_i's in terms of κ and τ appropriately to obtain a complete set of partial differential equaitons. This essentially means restricting the geometrical shape of the curve to certain specified forms. For example, with the specification

$$r = -q^*, \quad \zeta = -\tau_0/2 = \text{constant}, \tag{33a}$$

$$q = -(\kappa/2)\exp\left[-i\int (\tau - \tau_0)\, dx\right], \tag{33b}$$

and

$$A = -(i/2)\left[\omega_1 - \int^x (\tau - \tau_0)_t\, dx\right],$$

$$B = -C^* = -\tfrac{1}{2}(\omega_3 + i\omega_2)\exp\left(-i\int (\tau - \tau_0)\, dx\right) \tag{33c}$$

the AKNS system (1) is obtained.

In order to see these aspects more clearly, we can rewrite the scalar components of the Serret–Frenet equations (15) and the rigid body equations (22) in terms of the Dar-

boux vector

$$z_l = \frac{e_{2l} + ie_{3l}}{(1 - e_{1l})}, \quad e_{1l}^2 + e_{2l}^2 + e_{3l}^2 = 1 \quad (l = 1,2,3). \tag{34}$$

One obtains the following set of Riccati equations in place of (15) and (22):

$$z_{lx} = -i\tau z_l + (\kappa/2)(z_l^2 + 1), \tag{35}$$

$$z_{lt} = -i\omega_1 z_l + \tfrac{1}{2}(\omega_3 + i\omega_2)z_l^2 + \tfrac{1}{2}(\omega_3 - i\omega_2). \tag{36}$$

[Similar sets of equations are obtained for the cyclic combinations

$$\hat{z}_l = \frac{e_{3l} + ie_{1l}}{1 - e_{2l}}, \quad \tilde{z}_l = \frac{e_{1l} + ie_{2l}}{1 - e_{3l}},$$

the consequences of which are the same as with (35) and (36) and so we do not discuss them here.] With the substitution

$$z_l = \tilde{v}_2/\tilde{v}_1, \tag{37}$$

a *particular* solution of the system (35) and (36) is

$$\tilde{v}_{1x} = \frac{i}{2}\tau\tilde{v}_1 - \frac{\kappa}{2}\tilde{v}_2, \tag{38a}$$

$$\tilde{v}_{2x} = \frac{\kappa}{2}\tilde{v}_1 - \frac{i}{2}\tau\tilde{v}_2, \tag{38b}$$

and

$$\tilde{v}_{1t} = \tfrac{1}{2}i\omega_1\tilde{v}_1 - \tfrac{1}{2}(\omega_3 + i\omega_2)\tilde{v}_2, \tag{39a}$$

$$\tilde{v}_{2t} = \tfrac{1}{2}(\omega_3 - i\omega_2)\tilde{v}_1 - \tfrac{1}{2}\omega_1\cdot\tilde{v}_2. \tag{39b}$$

By making the substitution

$$\tilde{v}_1 = v_1 \exp\left[\frac{i}{2}\int(\tau - \tau_0)\cdot dx\right], \tag{40a}$$

$$\tilde{v}_2 = v_2 \exp\left[-\frac{i}{2}\int(\tau - \tau_0)\cdot dx\right], \tag{40b}$$

it is clear that the identification (33) results in the AKNS system (1) and so also for the eigenvalue problem (2) and (3).

Finally the problem of determining the quantities $\alpha, \beta,$ and γ from Eq. (32) is facilitated by the fact that the homogeneous part of the set of ordinary linear differential Eqs. (32) is identical to the scalar component form of the Serret–Frenet Eqs. (15). Denoting

$$\Gamma = (\alpha,\beta,\gamma)^T \quad \text{and} \quad \Delta = (0, -\omega_3, -\omega_2)^T \tag{41a}$$

the solution to Eq. (32) which is

$$\Gamma_x = \mathsf{F}\Gamma + \Delta, \tag{41b}$$

is given by (see Ref. 7)

$$\Gamma = \mathsf{E}\left[\mathsf{C} + \int_{-\infty}^{x} dx'\, \mathsf{E}^{-1}(x')\Delta(x')\right], \tag{42}$$

where C is a constant matrix and E is the solution of the homogeneous part i.e.,

$$\mathsf{E}_x = \mathsf{FE}, \tag{43}$$

analogous to the scalar form of Eqs. (15). Thus for E, a (3×3) matrix could be constructed by knowing the solution of the Serret–Frenet vectors. To solve for these vectors e_{il}, we can use the fact that in all the soliton possessing evolution equations, Bäcklund transformations connecting two different solutions through the solution of the Riccati equation can be constructed.[16] For example, in the case of the nonlinear Schrödinger equation[5,16]

$$iq_t + q_{xx} + 2|q|^2 q = 0, \tag{44a}$$

two solutions q and q' [see Eqs. (33)] are related by

$$q' - q = (\tau_0 + \tau_0^*)\cdot z_l/(|z_l|^2 - 1)\cdot\exp\left[-i\int(\tau - \tau_0)\,dx\right]. \tag{44b}$$

Similar results for all the other equations can be found in Ref. 16. Thus knowing q' and q, z_l can be constructed. Then it is a known result in differential geometry[7,17] that for a given particular solution of z_l and curvature κ and torsion τ, the curve may be referred to a rectangular coordinate system. Similar results for $\hat{z}_l$ and $\tilde{z}_l$ [discussed under Eq. (36)] hold. With these results the matrix E is constructed and hence the solution (42). Thus the complete dynamics of the string and rigid body can be determined.

Another way of solving the equation of motion for the vectors $\mathbf{e}_i$, is to put the time variation of these vectors as given by Eqs. (22) or (26) into the inverse scattering formalism. Note that they are already in the form of Lax's equation as discussed by Hermann[10] in connection with his Moyal–Bracket construction.[10] We illustrate the case of the nonlinear Schrödinger equation (44a) in the following:

Equation (44a) corresponds to the situation in which the angular velocity vector is of the form [18]

$$\omega_1 = (\kappa_{xx}/\kappa - \tau^2), \quad \omega_2 = -\kappa_x, \quad \omega_3 = -\kappa\tau. \tag{45}$$

Considering Eq. (21) for $i = 1$,

$$\mathbf{e}_{1t} = \boldsymbol{\omega}\times\mathbf{e}_1 = -\mathbf{e}_1''/2\times\mathbf{e}_1, \tag{46a}$$

as noted in Sec. 2 B, Eq. (26), could be put in the matrix form

$$\mathsf{e}_{1t} = [\mathsf{e}_1,\omega] \tag{46b}$$

with

$$\mathsf{e}_1^2 = 1, \quad \mathsf{e}_1^{\dagger} = -\mathsf{e}_1, \quad \mathrm{tr}\mathsf{e}_1 = 0, \quad \omega = -\mathsf{e}_{1xx}/2. \tag{46c}$$

Now defining the operators

$$\mathsf{L} = \mathsf{e}_1\frac{\partial}{\partial x}, \tag{47a}$$

$$\mathsf{B} = \mathsf{a}\frac{\partial}{\partial x} + \mathsf{b}\frac{\partial^2}{\partial x^2}, \tag{47b}$$

we form the Lax's operator equation

$$\frac{\partial\mathsf{L}}{\partial t} = [\mathsf{L},\mathsf{B}]. \tag{47c}$$

This results in the following condition for the unknown matrix coefficients a and b:

$$\mathsf{b} = \mathsf{e}_1\mathsf{b}\mathsf{e}_1, \tag{48a}$$

$$\mathsf{a} + \mathsf{b}_x = \mathsf{e}_1\mathsf{a}\mathsf{e}_1 + 2\mathsf{e}_1\mathsf{b}\mathsf{e}_{1x}, \tag{48b}$$

$$\mathsf{a}_x - \omega = \mathsf{e}_1\mathsf{a}\mathsf{e}_{1x} + \mathsf{e}_1\mathsf{b}\mathsf{e}_{1xx} - \mathsf{e}_1\omega\mathsf{e}_1. \tag{48c}$$

A solution of the set (48) is

$$\mathsf{a} = \mathsf{e}_{1x}, \quad \mathsf{b} = 2\mathsf{e}_1. \tag{49}$$

Thus (47) is in the form amenable to solution through IST. A similar equation occurs for the spin vector of the continuum Heisenberg ferromagnet system whose initial value problem has been solved recently by Takhtajan.[19] Ear-

lier we had identified [18a] this with the nonlinear Schrödinger equation. Once $\mathbf{e}_1$ is known, $\mathbf{e}_2$ and $\mathbf{e}_3$ can be determined with a knowledge of the curvature and torsion, the latter two quantities being given in terms of the solution of the nonlinear Schrödinger equation.

4. PROLONGATION STRUCTURE, CONNECTION FORMS, AND FIBRE BUNDLES

We proceed to show how the above results may be related to the abstract concepts of Wahlquist and Estabrook[5] and Hermann[6] such as pseudopotentials, prolongation structures and connection forms of certain fibre bundle structures. This enables one to give a simple geometrical and physical meaning of these quantities.

Firstly, we note that Eqs. (15) and (22) of the space curves and rigid body may be rewritten in the suggestive form:

$$\mathsf{y}_x = \mathsf{F}(\kappa,\tau)\,\mathsf{y} \equiv (\tau \mathsf{X}_1 + 0\cdot\mathsf{X}_2 + \kappa \mathsf{X}_3)\mathsf{y} \tag{50}$$

and

$$\mathsf{y}_t = \mathsf{G}(\omega_j)\,\mathsf{y} \equiv (\omega_1 \mathsf{X}_1 + \omega_2 \mathsf{X}_2 + \omega_3 \mathsf{X}_3)\mathsf{y}, \tag{51}$$

where

$$\mathsf{y} = (\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)^T, \tag{52}$$

and the matrices X_1, X_2, X_3 are the SO(3) generators as discussed in Sec. 2 B Eq. (25). In the language of exterior differential forms[5,20] system (50)–(51) may be equivalently written as the Pfaffian 1-form

$$\Omega = \mathsf{dy} - \mathsf{F}\mathsf{y}\,dx - \mathsf{G}\mathsf{y}\,dt. \tag{53}$$

Then the 2-form

$$\begin{aligned} d\Omega &= [(\mathsf{F}_t\mathsf{y} - \mathsf{F}\mathsf{y}_t) - (\mathsf{G}_x\mathsf{y} - \mathsf{G}\mathsf{y}_x)]\,dx \wedge dt \\ &= \{\mathsf{F}_t - \mathsf{G}_x - [\mathsf{F},\mathsf{G}]\}\mathsf{y}\,dx \wedge dt = 0, \end{aligned} \tag{54}$$

in view of Eqs. (31a) and (50) and (51). However Eq. (31a) itself could be written in the one-differential form

$$\Omega_0 = \mathsf{F}\,dx - \left[\mathsf{G} - \int [\mathsf{F},\mathsf{G}]dx\right]dt. \tag{55}$$

To this, an exact 1-differential dy may always be adjoined, i.e., the original Ω_0 given by (55) may be prolonged to a new

$$\Omega_0 = dy + \mathsf{F}\,dx - \left[\mathsf{G} - \int [\mathsf{F},\mathsf{G}]dx\right]dt \tag{56}$$

without disturbing the original set of Eqs. (31a). This is simply due to the identity $d(dy) = 0$. Note the underlying SO(3) structure of the matrices F and G in Ω_0.

Now the structure Ω as given by Eq. (53) can be considered as a further prolongation of the new Ω_0 as given by Eq. (56), in which the linear pseudopotential y is brought into the second and third coefficients on the right-hand side of (53). The nomenclature pseudopotential[5] is due to the fact that it cannot be now expressed as a simple quadrature. We also note that the underlying SO(3) Lie algebraic structure is still maintained. All these facts are exactly included in the phenomena discussed by Wahlquist and Estabrook (See Ref. 5. Secs. 2–4). In their procedure one searches for a suitable F and G having these properties. Here they follow in a natural way for the particular set of equations under consideration.

Similarly the set of Eqs. (38) and (39) for $\tilde{\mathsf{v}}$, where $\tilde{\mathsf{v}}^T = (\tilde{v}_1,\tilde{v}_2)$ can be written in the form

$$\tilde{\mathsf{v}}_x = \tilde{\mathsf{F}}\tilde{\mathsf{v}} \equiv (\tau\tilde{\mathsf{X}}_1 + 0\cdot\tilde{\mathsf{X}}_2 + \kappa\tilde{\mathsf{X}}_3)\tilde{\mathsf{v}}, \tag{57a}$$

$$\tilde{\mathsf{v}}_t = \tilde{\mathsf{G}}\tilde{\mathsf{v}} \equiv (\omega_1\tilde{\mathsf{X}}_1 + \omega_2\tilde{\mathsf{X}}_2 + \omega_3\tilde{\mathsf{X}}_3)\tilde{\mathsf{v}}, \tag{57b}$$

where now the (2×2) matrices $\tilde{\mathsf{X}}_1$, $\tilde{\mathsf{X}}_2$, $\tilde{\mathsf{X}}_3$ are the generators of the SU(2) group [homomorphic to SO(3)]. The linear pseudopotential $\tilde{\mathsf{v}}$ is obtained from y through a nonlinear coordinate transformation preserving the Lie algebraic structure and hence the content of the theory remains unchanged. A further one-parameter group of transformation (40) results in the eigenvalue problem (2) and (3).

Finally the set of Riccati equations (35) and (36) may be written as

$$z_{lx} = \hat{F}(\kappa,\tau,z_l)\,z_l = \left[-i\tau X_0 + \frac{\kappa}{2}X_1 - \frac{\kappa}{2}X_{-1}\right]z_l, \tag{58a}$$

$$\begin{aligned} z_{lt} &= \hat{G}(\omega_j,z_l)z_l \\ &= [-i\omega_1 X_0 + \tfrac{1}{2}(\omega_3 + i\omega_2)X_1 - \tfrac{1}{2}(\omega_3 - i\omega_2)X_{-1}]z_l. \end{aligned} \tag{58b}$$

Here the vectors

$$X_0 = z_l\frac{\partial}{\partial z_l}, \quad X_1 = z_l^2\frac{\partial}{\partial z_l}, \quad X_{-1} = \frac{-\partial}{\partial z_l} \tag{59}$$

are the generators of the SL(2,R) algebra [which is again homomorphic to SO(3) and SU(2)] satisfying the relations

$$[X_0,X_1] = X_1, \quad [X_0,X_{-1}] = -X_{-1}, \quad [X_1,X_{-1}] = 2X_0. \tag{60}$$

The z_l's are undoubtedly another equivalent class of pseudopotnetials as they were obtained through a coordinate transformation from the e_{il}'s. The prolongation structure associated with the quadratic pseudopotential z_l is then

$$\begin{aligned} \Omega_l = dz_l + &\left[\left(-\frac{\kappa}{2} + i\tau z_l - \frac{\kappa}{2}z_l^2\right)\right]dx \\ &+ [\tfrac{1}{2}(\omega_3 - i\omega_2) + i\omega_1 z_l - \tfrac{1}{2}(\omega_3 + i\omega_2)z_l^2]\,dt. \end{aligned} \tag{61}$$

These results then naturally lead to an identification of the basic fiber bundle structure involved. This can be best illustrated in conjunction with the results of Eqs. (58)–(61). A quadratic connection

$$\tilde{\omega} = \tilde{\omega}_0 + \tilde{\omega}_1 z_l + \tilde{\omega}_2 z_l^2 \tag{62}$$

may be defined, where

$$\tilde{\omega}_0 = \frac{\kappa}{2}dx + \tfrac{1}{2}(\omega_3 - i\omega_2)dt, \tag{63a}$$

$$\tilde{\omega}_1 = -i\tau\,dx - i\omega_1\,dt, \tag{63b}$$

$$\tilde{\omega}_2 = \frac{\kappa}{2}dx + \tfrac{1}{2}(\omega_3 + i\omega_2)dt. \tag{63c}$$

This induces the curvature forms

$$\theta_1 = d\tilde{\omega}_1 + 2\tilde{\omega}_0 \wedge \tilde{\omega}_2, \tag{64a}$$

$$\theta_2 = d\tilde{\omega}_2 + \tilde{\omega}_1 \wedge \tilde{\omega}_2, \tag{64b}$$

$$\theta_3 = d\tilde{\omega}_0 + \tilde{\omega}_0 \wedge \tilde{\omega}_1, \tag{64c}$$

having the basic SL(2,R) structure (60). One verifies that sectioning the forms, $\theta_i = 0$ ($i = 1,2,3$) leads to the original evolution equations (31a), by substituting (63) in (64).

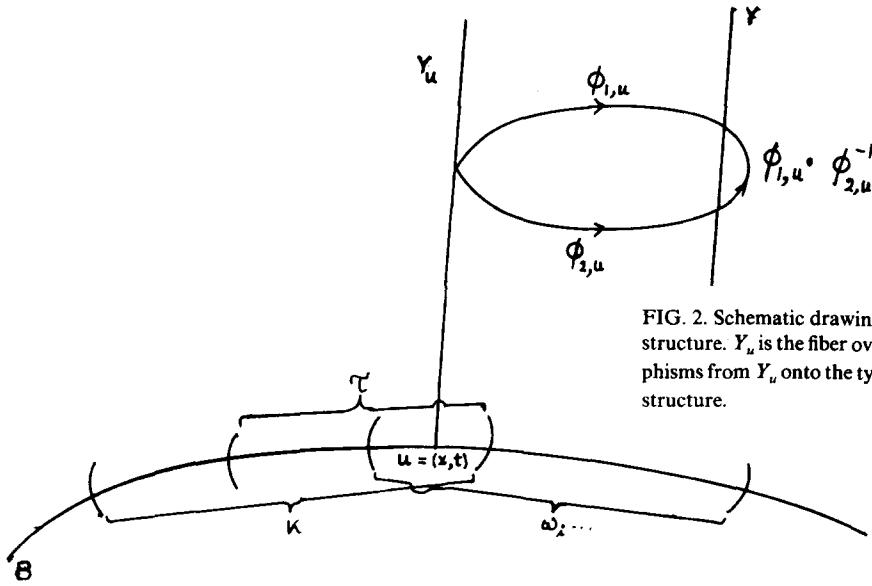

FIG. 2. Schematic drawing of the fiber bundle having the SL(2,R) group structure. Y_u is the fiber over $u = (x,t)$ and $\phi_{1,u}$ and $\phi_{2,u}$ are the homeomorphisms from Y_u onto the typical fiber Y with $\phi_{1,u}\cdot\phi_{2,u}^{-1}$ defining the SL(2,R) structure.

The underlying fiber bundle structure E is now apparent. It is the differential manifold having the collection of following objects:

(a) The base space B, covered by the family of coordinate neighborhoods $\{x,t,\kappa,\tau,\omega_i\}$;

(b) The one-dimensional space of z_l, called the fiber Y;

(c) The bundle space E which is locally the topological product of B and Y, i.e., $E = B \times Y$;

(d) The mapping π of E onto B equivalent to the original set of Eqs. (31a);

(e) The structural group SL(2,R) of the bundle that acts effectively and differentially on Y.

These facts are illustrated in Fig. 2.

5. DISCUSSION

Starting from an elementary consideration of the motion of an arbitrary rigid body constrained to a helical space curve and deriving the compatibility equations, we have obtained the class of nonlinear evolution equations solvable by the two-component inverse scattering phenomenology. The very nature of the phenomenon under consideration reveals the underlying group structures leading in a simple way to the associated concepts of pseudopotentials and prolongation structures. A natural fiber bundle structure also emerges thereby. It is thus possible to present a unified interpretation of this class of solitons possessing evolution equations through a mechanism, which is within the realm of an undergraduate laboratory experiment.

ACKNOWLEDGMENT

The author would like to express his warm thanks to Professor L.J.F. Broer for hospitality during his stay at the Eindhoven University of Technology, the Netherlands where a major part of this work was carried out.

[1]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[2]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[3]edited by R.M. Miura, *Bäcklund Transformations*, Lecture Notes in Mathematics (Springer, Berlin, 1974), Vol. 515.
[4]R. Bullogh and P. Caudrey, Eds., *Solitons* (Springer, Berlin, 1977).
[5]H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. **16**, 1 (1975); **17**, 1293 (1976).
[6]R. Hermann, Phys. Rev. Lett. **36**, 835 (1976).
[7]G.L. Lamb, Jr., Phys. Rev. Lett. **37**, 235 (1976); J. Math. Phys. **18**, 1654 (1977).
[8]C. Morris, J. Math. Phys. **18**, 533 (1977).
[9]J. Corones, (a) J. Math. Phys. **17**, 756 (1976); (b) J. Math. Phys. **18**, 533 (1977).
[10]R. Hermann, Phys. Rev. Lett. **37**, 1591 (1976).
[11]R.K. Dodd and J.D. Gibbon, Proc. R. Soc. London **358**, 287 (1978).
[12]F. Lund and T. Regge, Phys. Rev. D **14**, 1524 (1976).
[13]S. Kumei, J. Math. Phys. **16**, 2461 (1975); **18**, 256 (1977).
[14]L.J.F. Broer, J. Eng. Math. **4**, 195 (1970).
[15]J.J. Stoker, *Differential Geometry* (Wiley, New York, 1969), Chap. IV.
[16]H.H. Chen, Phys. Rev. Lett. **33**, 925 (1974).
[17]L.P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Dover, New York, 1960).
[18]M. Lakshmanan, Phys. Lett. A (a) **61**, 53 (1977); (b) **64**, 353 (1978).
[19]L.A. Takhtajan, Phys. Lett. A **64**, 235 (1978).
[20]Y. Choquet-Bruhat, C. Dewitt-Morette, and M.D. Dillard Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).

Remarks on a theorem by G. Epifanio, "On the matrix representation of unbounded operators" [J. Math. Phys., 17, 1688 (1976)][a)]

G. Epifanio
Istituto di Fisica dell'Università di Palermo, Palermo, Italy

C. Trapani
Istituto di Matematica dell'Università di Palermo, Palermo, Italy

(Received 18 January 1979)

We give a generalization of a theorem concerning the change of basis for the matrix representation of unbounded operators defined in a scalar product space. We introduce for the proof a suitable structure which can be useful when one has to make operations with operators defined between different scalar product spaces.

1. INTRODUCTION

In a previous paper,[1] one of us examined the problem of the matrix representation of the elements belonging to the *-algebra C_D of unbounded operators[2] in a scalar product space D. It was proved that the problem is solvable in a way analogous to that used for the bounded operators in a Hilbert space H.

On discussing the problem of the change of basis, the additional hypothesis that the operator associated with it be an automorphism of D was made. Only in this case, in fact, was it possible to know the nature of this operator and of the matrix associated with it. The given statement[3] holds true, however, in general, as one can easily see by simple formal matrix manipulations.[4]

It is obvious that if D is complete (Hilbert space) such a difficulty does not arise, because each operator associated with a change of basis in a Hilbert space is a unitary operator.

In this paper we prove that the mentioned theorem of change of basis can also be stated in the general case giving an exact meaning to all the operators and matrices; for this purpose we introduce some simple structures which can also be useful in similar cases and particularly when one has to deal with operators defined between different scalar product spaces and to make operations among them.

2. THE GENERAL THEOREM OF THE CHANGE OF BASIS

For an easier understanding of the text, we will recall some definitions and theorems given in a previous paper.[5]

Definition 1: Let D be a scalar product space. We say that a linear operator A defined on D has an adjoint A^* in D whenever there exists a linear operator A^* defined on D such that

$$\forall\, \varphi,\psi\in D \quad (A\varphi,\psi) = (\varphi,A^*\psi).$$

We call C_D the set of the linear operators that are defined on D and have an adjoint in D.

Theorem 1: For any scalar product space D, C_D (endowed with the natural operations) is a *-algebra of closed operators.

Theorem 2: In order that the operator A belong to C_D it is necessary and sufficient that the operator A be continuous in the weak topology $\sigma(D,D)$ determined in D by the set of seminorms

$$\{\varphi\to|(\varphi,\psi)| \quad |\ \psi\in D\}$$

Let us introduce now the space $C_{D,D'}$.

Definition 2: Let D and D' be two scalar product spaces; we say that a linear operator A from D to D', defined everywhere in D, has an adjoint A^* from D' to D, if there exists an operator A^*, defined everywhere in D', such that

$$\forall\varphi\in D,\ \forall\varphi'\in D', \quad (A\varphi,\varphi') = (\varphi,A^*\varphi').$$

We call $C_{D,D'}$ the set of all linear operators from D into D' and which have an adjoint from D' into D.

It is easily seen that $C_{D,D'}$, endowed with the natural operations is a linear space of closed operators and that the map $A\in C_{D,D'}\to A^*\in C_{D',D}$ provides an anti-isomorphism between the spaces $C_{D,D'}$ and $C_{D',D}$. As an application of a known theorem[6] we have

Theorem 3: In order that the operator A belong to $C_{D,D'}$ it is necessary and sufficient that the operator A be continuous for the weak topologies $\sigma(D,D)\to\sigma(D',D')$.

We shall now discuss the problem of the matrix representation of the operators of $C_{D,D'}$. We omit the proof of some theorems because they are analogous to those proved[7] for C_D. In the sequel we assume that the scalar product spaces that we consider are separable.

Definition 3: Let A be a linear operator from the scalar product space D into the scalar product space D', let (e_ν) and (e'_ν) be orthonormal bases, respectively, in D and in D' and $M = (A_{\mu\nu})$ an infinite matrix. We say that the matrix M represents the operator A with respect to the bases (e_ν) and (e'_ν) if $\forall\ \varphi = \Sigma_{\nu=1}^{\infty}\xi_\nu e_\nu\in D$ for $\varphi' = A\varphi$ with $\varphi' = \Sigma_{\nu=1}^{\infty}\xi'_\nu e'_\nu \in D'$ we have

$$\xi'_\mu = \sum_{\nu=1}^{\infty} A_{\mu\nu}\xi_\nu.$$

Using the linearity and the continuity of the operator $A\in C_{D,D'}$ stated in Theorem 3, the following proposition is easily proved.

Theorem 4: Every operator $A\in C_{D,D'}$ admits a matrix representation with respect to any orthonormal basis (e_ν) in D and any orthonormal basis (e'_ν) in D'. The matrix $(A_{\mu\nu})$ is defined by the relations

$$A_{\mu\nu} = (Ae_\nu,e'_\mu), \quad \nu,\mu = 1,2,\cdots.$$

[a)]Work supported by C.R.R.N.S.M. Istituto di Fisica, Università di Palermo, Palermo, Italy.

 0022-2488/79/081673-03$01.00

In this representation to the operator $A^* \in C_{D',D}$ we associate the matrix

$$M^* = (A^*_{\mu\nu}) \quad \text{with} \quad A^*_{\mu\nu} = \bar{A}_{\nu\mu}.$$

Theorem 5: Let D and D' be two scalar product spaces, (e_ν) and (e'_ν) orthonormal bases in D and in D', respectively, d and d' the linear manifolds of l^2 canonically isomorphic to D and D' respectively. In order that the matrix $M = (A_{\mu\nu})$ represent an operator $A \in C_{D,D'}$ it is necessary and sufficient that

(a) if $(\xi_\nu)_{\nu\in\mathbb{N}} \in d$, then $\left(\sum_{\nu=1}^{\infty} A_{\mu\nu}\xi_\nu\right)_{\mu\in\mathbb{N}} \in d'$,

(b) if $(\xi'_\mu)_{\mu\in\mathbb{N}} \in d'$, then $\left(\sum_{\mu=1}^{\infty} \bar{A}_{\mu\nu}\xi'_\mu\right)_{\nu\in\mathbb{N}} \in d$,

(c) if $(\xi_\nu)_{\nu\in\mathbb{N}} \in d$, and $(\xi'_\mu)_{\mu\in\mathbb{N}} \in d'$, then

$$\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty} \xi_\nu A_{\mu\nu} \bar{\xi}'_\mu = \sum_{\nu=1}^{\infty}\sum_{\mu=1}^{\infty} \xi_\nu A_{\mu\nu} \bar{\xi}'_\mu.$$

Besides, if we call $\mathcal{M}_{d,d'}$ the set of matrices satisfying the conditions (a), (b), (c) and if we define the matrix operations as usual, $\mathcal{M}_{d,d'}$ is a linear space and the map $A \in C_{D,D'} \to M(A) \in \mathcal{M}_{d,d'}$ provides an isomorphism of $C_{D,D'}$ onto $\mathcal{M}_{d,d'}$.

In particular if $D = D'$ and $(e_\nu) = (e'_\nu)$, then $\mathcal{M}_{d,d'} = \mathcal{M}_d$ is a *-algebra and the isomorphism between C_D and $\mathcal{M}_d$ is a *-isomorphism of *-algebras.[8]

Let us introduce a structure which allows us to prove some propositions in a very simple way.

If D and D' are two scalar product spaces, we consider the scalar product space $D \oplus D'$, direct sum of D and D'; operations and scalar product are defined in the usual way.

It is now possible to introduce the *-algebra $C_{D\oplus D'}$ of linear operators in $D \oplus D'$ which have an adjoint in $D \oplus D'$. It is easily seen that every operator $A \in C_{D\oplus D'}$ may be written as a matrix of the type

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where $A_{11} \in C_D$, $A_{12} \in C_{D',D}$, $A_{21} \in C_{D,D'}$, and $A_{22} \in C_{D'}$.

We remark that to an operator $A \in C_D$ we may associate an operator $\hat{A} \in C_{D\oplus D'}$, such that $\hat{A} \mid D = A$, but the operator $\hat{A}$ is not uniquely determined; we choose the following extension

$$\hat{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}.$$

In an analogous way we can associate an operator of $C_{D\oplus D'}$ to an operator of $C_{D,D'}$, etc.

It is evident that these extensions provide an isomorphism of C_D and $C_{D'}$ onto two subalgebras of $C_{D\oplus D'}$ and an isomorphism of $C_{D,D'}$ and $C_{D',D}$ onto two subspaces of $C_{D\oplus D'}$.

If (e_ν) is a basis in D and (e'_ν) is a basis in D', then the sequence $(e_\nu \oplus 0, 0 \oplus e'_\mu)_{\nu,\mu\in\mathbb{N}}$ is a basis for $D \oplus D'$, hence if d and d' are the linear manifolds of l^2 which are canonically isomorphic to D and D', the matrix representation of the operators of $C_{D\oplus D'}$ with respect to the basis $(e_\nu \oplus 0, 0 \oplus e'_\mu)_{\mu,\nu\in\mathbb{N}}$ provides a *-algebraic isomorphism of $C_{D\oplus D'}$ onto $\mathcal{M}_{d\oplus d'}$.

The introduction of the *-algebras $C_{D\oplus D'}$ and $\mathcal{M}_{d\oplus d'}$ allows us to "quietly" use the matrices and to make operations among them, because their nature is well determined; in fact, it is known that generally these operations must be made very carefully, for instance, the multiplication of infinite matrices can also be not associative.

Theorem 6: Let A be an operator of C_D and B an operator of $C_{D,D'}$; let $\mathcal{M}_d$ and $\mathcal{M}_{d,d'}$ be the *-algebra and the linear space of matrices isomorphic, respectively, to C_D and $C_{D,D'}$ for the choice of the basis (e_ν) in D and (e'_ν) in D'. Then the operator BA belongs to $C_{D,D'}$ and for the matrices $M(B)$, $M(A)$ and $M(BA)$ representing the operators B, A and BA, respectively, in $\mathcal{M}_{d,d'}$, $\mathcal{M}_d$ and $\mathcal{M}_{d,d'}$ the following relation is valid:

$$M(BA) = M(B)\,M(A).$$

Analogously, if $A \in C_{D,D'}$ and $B \in C_{D',D}$ then $AB \in C_{D'}$ and

$$M(AB) = M(A)\,M(B).$$

Proof: We remark only that the above relations are justified by the fact that $\mathcal{M}_d$, $\mathcal{M}_{d'}$, and $\mathcal{M}_{d,d'}$ are isomorphic respectively to subalgebras and to a subspace of the *-algebra $\mathcal{M}_{d\oplus d'}$.

The definitions and theorems that we have given above supply a simple way to prove the theorem mentioned[3] in the Introduction which we recall for the reader's convenience.

Theorem 7: Let (e_ν) and (e'_ν) be two orthonormal bases in the scalar product space D, $M(A)$ and $M'(A)$ the matrices representing any operator $A \in C_D$ with respect to the bases (e_ν) and (e'_ν), respectively. Let U be the operator associated with the change of basis, that is $Ue_\nu = e'_\nu$; if U is an automorphism of D, then the following relation is valid

$$M'(A) = M(U^{-1})\,M(A)\,M(U),$$

where $(M(U))_{\mu\nu} = (Ue_\nu, e_\mu)$.

The hypothesis that U be an automorphism of D was added because the space D being not complete, the operator U associated with the change of basis has not, generally, the whole space as domain. In other words, if we consider D as a dense linear manifold of its norm completion $\hat{D}$, the linear manifold $D' = UD$ is generally, different from D, but $D \cap D'$ does not reduce only to $\{0\}$ [both the bases (e_ν) and (e'_ν), in fact, belong to it]. Notice that if we call d the linear manifold of l^2 canonically isomorphic to D for the choice of the basis (e_ν), D' contains all vectors of $\hat{D}$ of the type $\Sigma\lambda_\nu e'_\nu$ with $(\lambda_\nu)_{\nu\in\mathbb{N}} \in d$ and the map $\varphi = \Sigma_{\nu=1}^{\infty} \lambda_\nu e_\nu \to \varphi' = \Sigma_{\nu=1}^{\infty} \lambda_\nu e'_\nu$ provides an isomorphism between D and D' and so the map $A \in C_D \to UAU^{-1} \in C_{D'}$ provides an isomorphism between C_D and $C_{D'}$.

Here the operator U is well identified as a unitary operator from D to D' and one can also characterize the associated matrix.

Lemma 8: Let D be a scalar product space, (e_ν) a basis in D. Let U be the isometric operator in D which changes the basis (e_ν) into the basis (e'_ν) of D. If we call D' the space UD, then the operators $A \in C_D$ and $UAU^{-1} \in C_{D'}$ are represented, the first with respect to the basis (e_ν) and the second with respect to the basis (e'_ν) by the same matrix.

Proof: We indicate with M' the matrix which represents

an operator of $C_{D'}$ with respect to the basis (e'_ν) and with M the matrix which represents an operator of C_D with respect to the basis (e_ν). Because A is an operator in D and U an operator from D to D', the thesis of the theorem follows easily from the relations below:

$$(M'(UAU^{-1}))_{\mu\nu} = (UAU^{-1}e'_\nu, e'_\mu) = (AU^{-1}e'_\nu, U^{-1}e'_\mu)$$
$$= (Ae_\nu, e_\mu) = (M(A))_{\mu\nu}.$$

We conclude now the discussion of the problem of the change of basis, giving the generalization of the above theorem to the case in which the operator associated with the change of basis is not necessarily an automorphism of D.

Theorem 9: Let (e_ν) and (e'_ν) be two orthonormal bases in the scalar product space D, $M(A)$ and $M'(A)$ the matrices representing any operator $A \in C_D$ with respect to the bases (e_ν) and (e'_ν) respectively. Let U be the operator associated with the change of basis, that is $Ue_\nu = e'_\nu$, then the following relation is valid:

$$M'(A) = M(U^{-1})\,M(A)\,M(U),$$

where $(M(U))_{\mu\nu} = (Ue_\nu, e_\mu)$.

Proof: By Lemma 8 and Theorem 6

$$M(A) = M(UAU^{-1}) = M'(U)\,M'(A)\,M'(U^{-1}),$$

where

$$(M'(U))_{\mu\nu} = (Ue'_\nu, e'_\mu) = (Ue_\nu, e_\mu) = (M(U))_{\mu\nu},$$
$$(M'(U^{-1}))_{\mu\nu} = (U^{-1}e'_\nu, e'_\mu) = (U^{-1}e_\nu, e_\mu)$$
$$= (M(U^{-1}))_{\mu\nu},$$

hence

$$M'(A) = M'(U^{-1})M(A)M'(U)$$
$$= M(U^{-1})M(A)M(U).$$

ACKNOWLEDGMENT

The authors thank Dr. G. Riela for stimulating discussions.

[1] G. Epifanio, J. Math. Phys. **17**, 1688 (1976).
[2] See Definition 1.
[3] See Ref. 1, Theorem 6.
[4] See, for instance, M.H. Stone, *Linear Transformations in Hilbert Space* (Am. Math. Soc., New York, 1966), Chapter III, Theorem 3.1.
[5] See Ref. 1 and also R. Ascoli, G. Epifanio, and A. Restivo, Riv. Mat. Univ. Parma **3**, 21 (1974).
[6] N. Bourbaki, *Elements de Mathématique*: *Espaces Vectoriels Topologiques* (Hermann, Paris, 1965), 2nd ed., Chapter II, §6, Proposition 5 and corollary.
[7] See Ref. 1.
[8] See Ref. 1, Theorems 4 and 5. Notice that if D and D' are different but isomorphic scalar product spaces and the unitary operator U which realizes the isomorphism of D onto D' maps the basis (e_ν) into (e'_ν), then also in this case $\mathcal{M}_{d,d'} = \mathcal{M}_d$ but the isomorphism between $C_{D,D'}$ and $\mathcal{M}_d$ is an isomorphism of vector spaces.

Transformation between the normal and antinormal expansions of boson operators

Dan Shalitin and Yoel Tikochinsky

Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel
(Received 1 August 1978)

We show that the expansion coefficients for the normal and antinormal forms of a boson operator are related by a simple transform.

1. INTRODUCTION

Among all possible expansions of an analytic boson operator two forms are particularly useful—the normal and antinormal expansion. Thus, given an operator

$$f(a,a^\dagger)=\sum c_{\cdots ijkl\cdots}a^i a^{\dagger j}a^{\dagger k}a^l\cdots,$$

one can use the commutation relation $[a,a^\dagger]=1$ to bring all the creation operators to the left of the destruction operators. This results in the [unique] normal expansion

$$f=f^{(n)}(a,a^\dagger)\equiv\sum_{r,s}f^{(n)}_{rs}a^{\dagger r}a^s. \tag{1}$$

Similarly, one can obtain the (unique) antinormal expansion

$$f=f^{(a)}(a,a^\dagger)\equiv\sum_{r,s}f^{(a)}_{rs}a^r a^{\dagger s}. \tag{2}$$

The two forms arise naturally in the coherent state representation.[1] Thus, the diagonal elements of f in this representation are given by the normal form

$$\langle\alpha|f|\alpha\rangle=f^{(n)}(\alpha,\alpha^*), \tag{3}$$

and its projection-operator expansion by the antinormal form

$$f=\int\frac{d^2\alpha}{\pi}f^{(a)}(\alpha,\alpha^*)\,|\,\alpha\rangle\langle\alpha\,|\,. \tag{4}$$

Again, the expectation value of f, in a state characterized by a density operator ρ, is given by[2]

$$\langle f\rangle=\mathrm{tr}(\rho f)=\int\frac{d^2\alpha}{\pi}\rho^{(a)}(\alpha,\alpha^*)f^{(n)}(\alpha,\alpha^*)$$

$$=\int\frac{d^2\alpha}{\pi}\rho^{(n)}(\alpha,\alpha^*)f^{(a)}(\alpha,\alpha^*)$$

$$=\int\frac{d^2\alpha}{\pi}\langle\alpha|\rho|\alpha\rangle f^{(a)}(\alpha,\alpha^*). \tag{5}$$

The last line expresses $\langle f\rangle$ in a classiclike form—a subject of a recent publication.[3] Thus, according to Eqs. (2) and (5), to every analytic operator $f(a,a^\dagger)$ corresponds a unique "classical" function $f^{(a)}(\alpha,\alpha^*)$ the integral of which over phase space (Reα,Imα) with a positive weight $\langle\alpha|\rho|\alpha\rangle$, gives the quantum expectation value $\langle f\rangle$.

Having motivated our interest in the normal and antinormal forms of the boson operators, we shall now state our result[3]

$$f^{(n)}_{l!m!}=[l!m!]^{-1}\sum_{k=0}^{\infty}f^{(a)}_{k+m,k+l}(k+l)!(k+m)!/k!, \tag{6a}$$

$$f^{(a)}_{l!m!}=[l!m!]^{-1}\sum_{k=0}^{\infty}(-1)^k f^{(n)}_{k+m,k+l}(k+l)!(k+m)!/k!. \tag{6b}$$

Section 2 is devoted to the derivation of Eq. (6). The general problem of convergence is not discussed. We assume that all series involved converge, and that the order of summations may be changed at will. This is trivially true for every finite expansion. The series (6) may diverge but still be useful. For example, by embedding the operator f in a one-parameter family of analytic operators $f(\lambda)$, we may evaluate the sum in Eq. (6) for λ within the range of convergence. The appropriate form of f is then deduced by the assumed analyticity. We shall encounter such an example in Sec. 3.

2. DERIVATION OF THE NORMAL–ANTINORMAL TRANSFORM

Our starting point is the diagonal matrix element of Eq. (4), namely,

$$f^{(n)}(\beta,\beta^*)=\int\frac{d^2\alpha}{\pi}f^{(a)}(\alpha,\alpha^*)e^{-|\alpha-\beta|^2}. \tag{7}$$

Using the identity[4]

$$\int\frac{d^2\alpha}{\pi}\exp(-|\alpha|^2+\beta^*\alpha)\alpha^r g(\alpha^*)=\left(\frac{\partial}{\partial\beta^*}\right)^r g(\beta^*), \tag{8}$$

valid for every analytic function $g(z)$, we can perform the integration in Eq. (7) to obtain

$$\sum_{r,s}f^{(n)}_{rs}\beta^{*r}\beta^s$$

$$=\sum_{r,s}f^{(a)}_{rs}\exp(-|\beta|^2)\left(\frac{\partial}{\partial\beta^*}\right)^r[\exp(\beta^*\beta)\beta^{*s}]$$

$$=\sum_{r,s}f^{(a)}_{rs}\sum_{k=0}^{[r,s]}\binom{r}{k}\binom{s}{k}k!\beta^{(r-k)}\beta^{*(s-k)}. \tag{9}$$

Here, $[r,s]=\min(r,s)$. Since the diagonal element $f^{(n)}(\beta,\beta^*)$ determines f completely, we secure

$$\sum_{r,s}f^{(n)}_{rs}a^{\dagger r}a^s=\sum_{r,s}f^{(a)}_{rs}\sum_{k=0}^{[r,s]}\binom{r}{k}\binom{s}{k}k!a^{\dagger(s-k)}a^{(r-k)}$$

$$=\sum_{r,s}f^{(a)}_{rs}a^r a^{\dagger s}, \tag{10}$$

 0022-2488/79/081676-03$01.00

where the last equality follows from Eqs. (1) and (2). In particular, we have

$$a^r a^{\dagger s} = \sum_{k=0}^{[r,s]} \binom{r}{k}\binom{s}{k} k!\, a^{\dagger(s-k)} a^{(r-k)}. \tag{11}$$

Note that the validity of Eq. (11) depends solely on the commutation relation $[a,a^\dagger] = 1$.

We would like to invert the last relation, that is, to expand the normal monomial in terms of the antinormal monomials. To this end, consider the operators

$$b = a^\dagger, \quad \hat{b} = -a. \tag{12}$$

Substituting in Eq. (11), we have

$$(-1)^r \hat{b}^r b^s = \sum_{k=0}^{[r,s]} \binom{r}{k}\binom{s}{k} k!\, b^{(s-k)} \hat{b}^{(r-k)} (-1)^{r-k}. \tag{13}$$

Since $[b,\hat{b}] = 1$ [and Eq. (13) is valid for any pair of operators satisfying the boson commutation relations], we secure[5]

$$a^{\dagger r} a^s = \sum_{k=0}^{[r,s]} (-1)^k \binom{r}{k}\binom{s}{k} k!\, a^{(s-k)} a^{\dagger(r-k)}. \tag{14}$$

We note in passing, that Eqs. (11) and (14) allow the calculation of the commutator $[a^r,a^{\dagger s}]$ and the anticommutator $\{a^r,a^{\dagger s}\}$ recursively in terms of lower commutators and anticommutators.

Returning to Eq. (10), we shall rearrange the summations in the middle term as follows,

$$\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\sum_{k=0}^{[r,s]} g(r,s,k) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} g(m+k,l+k,k). \tag{15}$$

Hence,

$$\sum_{l,m} f_{lm}^{(n)} a^{\dagger l} a^m = \sum_{l,m}\left[\sum_k f_{m+k,l+k}^{(a)} \binom{m+k}{k}\binom{l+k}{k} k!\right] a^{\dagger l} a^m. \tag{16}$$

Equating the coefficients of $a^{\dagger l}a^m$ on both sides, we secure our first result, Eq. (6a). To obtain the second result, multiply Eq. (14) by $f_{rs}^{(n)}$ and sum over r and s. Rearranging the resulting summations on the right-hand side as in Eq. (15), we find,

$$\sum_{r,s} f_{rs}^{(n)} a^{\dagger r} a^s = \sum_{l,m}\left[\sum_k (-1)^k f_{m+k,l+k}^{(n)} \binom{m+k}{k}\binom{l+k}{k} k!\right] a^l a^{\dagger m} = \sum_{l,m} f_{lm}^{(a)} a^l a^{\dagger m}. \tag{17}$$

Equation (6b) then follows from the last equality.

3. EXAMPLE

In order to illustrate our result, we shall work out the following example. The normal expansion of $e^{-\lambda a^\dagger a}$ is given by[6]

$$e^{-\lambda a^\dagger a} = \sum_l \frac{(e^{-\lambda}-1)^l}{l!} a^{\dagger l} a^l. \tag{18}$$

We shall obtain the antinormal expansion for this operator[7] in two ways: (a) by performing the summation in Eq. (6b), and (b), by applying the transformation (12) directly to $e^{-\lambda a^\dagger a}$.

(a) By Eqs. (18) and (6b) the only nonvanishing antinormal coefficients are

$$f_{ll}^{(a)} \equiv f_l^{(a)} = [l!]^{-2} \sum_{k=0}^{\infty} (-1)^k (e^{-\lambda}-1)^{k+l} (k+l)!/k!. \tag{19}$$

In particular,

$$f_0^{(a)} = \sum (-1)^k (e^{-\lambda}-1)^k = e^\lambda \quad \text{for } |e^{-\lambda}-1| < 1. \tag{20}$$

It is easy to establish a recurrence relation for $f_l^{(a)}$:

$$f_{l+1}^{(a)} = [(l+1)!]^{-2} \sum_k (-1)^k (e^{-\lambda}-1)^{k+l+1} \times \left[\frac{(k+l)!}{k!}\right](k+l+1) = [(l+1)]^{-1}(e^{-\lambda}-1) f_l^{(a)} - (e^{-\lambda}-1) f_{l+1}^{(a)}.$$

Hence

$$f_{l+1}^{(a)} = [(l+1)]^{-1}(1-e^\lambda) f_l^{(a)}. \tag{21}$$

The solution of this recurrence relation with the initial condition (20), is

$$f_{ll}^{(a)} = e^\lambda (1-e^\lambda)^l / l!. \tag{22}$$

Since $e^{-\lambda a^\dagger a}$ is an analytic function of λ, this result is valid for all values of λ.

(b) Applying the transformation (12) to Eq. (18), we have

$$e^{\lambda b\hat{b}} = \sum_l \frac{(e^{-\lambda}-1)^l}{l!} (-1)^l b^l \hat{b}^l = e^{\lambda(1+\hat{b}b)} = e^\lambda e^{\lambda \hat{b}b}.$$

Hence

$$e^{\lambda \hat{b}b} = e^{-\lambda} \sum_l \frac{(1-e^{-\lambda})^l}{l!} b^l \hat{b}^l. \tag{23}$$

Since the last relation holds for any pair of boson operators, we obtain (replacing also λ by $-\lambda$)

$$e^{-\lambda a^\dagger a} = e^\lambda \sum_l \frac{(1-e^\lambda)^l}{l!} a^l a^{\dagger l}. \tag{24}$$

[1] W.H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973).

[2]Reference 1, p. 147.

[3]A. Lonke, J. Math. Phys. **19**, 1110 (1978). The "classical" representation [Eq. (5)] was apparently not known to Lonke. By a lengthy and mathematically inadmissible way, Lonke arrives at his central result [Eq. (3.10) of his paper] which *is* correct. In view of Eq. (5), Lonke's equation (3.10) is equivalent to Eq. (6b) of the present paper. Indeed, it was the simplicity of his result which encouraged us to find a simple way to establish the normal–antinormal transform (6a) and (6b).

[4]See, for example, Ref. 3.

[5]An alternative way to invert Eq. (11) is to multiply Eq. (9) by $e^{il\varphi}$, where $\beta = Re^{i\varphi}$, integrate over φ and replace in the resulting equation R^2 by $-x$. After the replacement, the right-hand side becomes a linear combination of Laguerre polynomials $L_n^l(x)$, and their orthogonality property can be invoked to accomplish the inversion.

[6]Reference 1, p. 156.

[7]Reference 1, p. 159.

SSS05157 JMAPA,00DEC878/12 – 10

Markov fields in noncommutative probability theory on W^* algebras

G. O. S. Ekhaguere

Department of Mathematics, University of Ibadan, Ibadan, Nigeria
(Received 8 June 1978)

Many important results in ordinary probability theory do not have extensions in noncommutative probability theory. Here we consider Markov fields, and we prove that new Markov fields may be generated from old ones by means of multiplicative measurable operators. There is an analog of this result in ordinary probability theory involving multiplicative functionals. It is envisaged that the result established here would be of use in the study of interacting Fermion quantum fields.

1. INTRODUCTION

A noncommutative theory of integration on W^*-algebras, in which the measure is required to be unitarily invariant and hence central, has been developed by Segal and his school[1-5] and the theory has had important applications to, for example, the theory of representations of locally compact groups,[4] the problem of duality for unimodular groups,[5] and the theory of relativistic quantum fields.[6-8] A new approach to noncommutative integration, based on Stinespring's notion[5] of *convergence in measure* of measurable operators, and which still requires the measure to be central, has recently been presented by Nelson.[2]

An earlier version of noncommutative integration than Segal's is due to Dye[10] who does not require the measure to be central. However, Dye requires the W^*-algebra on which integration is performed to be *finite* and *countably decomposable*. A W^*-algebra $\mathscr{B}$ is countably decomposable if any collection of mutually orthogonal projections in $\mathscr{B}$ is at most countable and $\mathscr{B}$ is finite if its identity if finite.[11] It is known[11] that any W^*-algebra of operators on a separable Hilbert space is countably decomposable.

Segal's and Dye's versions of noncommutative integration have their respective limitations; we refer to Ref. 12 for a critique of both versions. In Ref. 12, Gudder and Marchand outline a theory of noncommutative integration which does not require the sort of assumptions employed by Segal and Dye.

In this paper, we employ the Gudder–Marchand formulation of noncommutative integration in our study of *noncommutative Markov fields*. We prove a theorem which asserts that new Markov fields may be obtained from old ones by means of *multiplicative measurable operators*. This result is a generalization of the results in the theory of Markov processes which states roughly that new Markov processes may be obtained from old ones by means of multiplicative functionals.[13] Our result has applications in constructive quantum field theory[14] where Markov fields are curently playing a role of no small significance in the study of quantum fields.[15,16]

2. CONDITIONAL EXPECTATION

In this section, we present, for completeness, a brief summary of the basic ideas of noncommutative integration theory on a W^*-algebra as developed by Gudder and Marchand[11] (see also Ref. 17) and in the process, we establish our notation and terminologies.

In the sequel, $\mathscr{B}$ will denote a W^*-algebra, with identity 1, of operators on a Hilbert space Γ. We say that Γ is the underlying Hilbert space for $\mathscr{B}$. We denote the commutant of $\mathscr{B}$ by $\mathscr{B}'$ and write Z for the center $\mathscr{B}\cap\mathscr{B}'$ of $\mathscr{B}$. In almost any theory of noncommutative integration, the notion of *measurable operators*, due to Segal,[1] is of fundamental usefulness. We introduce this notion and other relevant concepts below.

2.1 *Definition*: A closed, densely defined linear operator $\mathbf{t}$ on Γ is said to be affiliated to $\mathscr{B}$ if $\mathbf{vt}\subseteq\mathbf{tv}$ for every unitary $\mathbf{v}$ in $\mathscr{B}'$. If $\mathbf{t}$ is affiliated to $\mathscr{B}$, we shall write $\mathbf{t}\eta\mathscr{B}$.

A linear set Λ in Γ is said to be associated with $\mathscr{B}$ (symbolically $\Lambda\eta\mathscr{B}$) if $\mathbf{v}(\Lambda)\subset\Lambda$ for every unitary $\mathbf{v}$ in $\mathscr{B}'$.

2.2 *Remark*: We note that if $\mathbf{t}$ is a bounded linear operator on Γ and $\mathbf{t}\eta\mathscr{B}$ then, by the double commutant theorem, $\mathbf{t}\in\mathscr{B}$.

2.3 *Definition*: Let Λ be a linear subset of Γ. Then Λ is said to be *strongly dense* in Γ with respect to $\mathscr{B}$ if

(i) $\Lambda\eta\mathscr{B}$

(ii) there is a sequence $\{\Lambda_n\}$ of subspaces of Γ, with $\Lambda_n\eta\mathscr{B}$ such that $\Lambda_n\subset\Lambda$, and

(iii) the projection operator of Γ onto the orthogonal complement $\Lambda_n^{\perp}$ of Λ_n is a finite projection in $\mathscr{B}$ and $\Lambda_n^{\perp}\downarrow 0$. We say that $\{\Lambda_n\}$ *defines* Λ.

2.4 *Definition*: An operator $\mathbf{t}$ on Γ is said to be *measurable* with respect to $\mathscr{B}$ provided that

(i) $t\eta\mathscr{B}$ and

(ii) $\mathbf{t}$ has a strongly dense domain.

2.5 *Remark*: (i) If $\mathbf{s}$ and $\mathbf{t}$ are measurable with respect to $\mathscr{B}$, then so are $\mathbf{s}^*$, $\mathbf{s}+\mathbf{t}$ and $\mathbf{st}$, where the strong sum and the strong product are employed here and $*$ denotes the adjoint operation.

(ii) In what follows, if $\mathscr{C}$ is a W^*-algebra of operators on Γ, let $p(\mathscr{C})$ denote the collection of all self-adjoint projections in $\mathscr{C}$.

2.6 *Definition*: A measure τ on $(\Gamma,\mathscr{B})$ is a nonnegative mapping

 0022-2488/79/081679-05$01.00

$\tau: p(\mathscr{B}) \to R_+ = [0, \infty)$

such that

(i) $\tau(0) = 0$ and

(ii) $\tau(\Sigma \mathbf{p}_n) = \Sigma \tau(\mathbf{p}_n)$, for any countable set $\{\mathbf{p}_n\}$ of mutually orthogonal projections in $p(\mathscr{B})$.

An *integral* τ on $(\Gamma, \mathscr{B})$ is a *faithful*, nonnegative linear functional

$\tau: \mathscr{B} \to \mathbb{C}$, the complex numbers

such that the restriction of τ to $p(\mathscr{B})$ is a measure. Thus an integral is a faithful normal nonnegative functional on $\mathscr{B}$ since normality is equivalent to countable additivity, i.e., condition (ii) in the definition of a measure.

An integral τ on $(\Gamma, \mathscr{B})$ is a *state* if $\tau(1) = 1$.

2.7 *Remark*: We note that we do not require here that τ be central as is usually assumed in Segal's approach to noncommutative integration.[1]

2.8 *Definition*: Let τ be a state on $(\Gamma, \mathscr{B})$. Then the triple $(\Gamma, \mathscr{B}, \tau)$ is called a (*noncommutative*) *probability space.*

If τ is central $(\Gamma, \mathscr{B}, \tau)$ is called a *probability gauge space* in the terminology of Segal.

2.9 *Definition*: Let $(\Gamma, \mathscr{B}, \tau)$ be a probability space. Then a sequence $\{\mathbf{a}_n\}$ of measurable operators is said to *converge in measure* to a measurable operator $\mathbf{a}$ if given $\delta > 0$, there is a sequence $\{\mathbf{p}_n\}$ of projections in $p(\mathscr{B})$ such that $\|(\mathbf{a}_n - \mathbf{a}) \times \mathbf{p}_n\| < \delta$ and $\tau(1 - \mathbf{p}_n) \to 0$, where $\| \cdot \|$ is the operator norm of $\mathscr{B}$.

2.10 *Remark*: In the sequel $L^0(\Gamma, \mathscr{B}, \tau)$ will denote the collection of all closed densely defined linear operators on Γ which are measurable with respect to $\mathscr{B}$; $L^0(\Gamma, \mathscr{B}, \tau)$ is a $*$ algebra, with the adjoint operation for its involution, and where the sum and product of operators are the strong sum and strong product, respectively. We assume in the following that $L^0(\Gamma, \mathscr{B}, \tau)$ is equipped with the topology of convergence in measure.

Next, for $\mathbf{a} \in L^0(\Gamma, \mathscr{B}, \tau)$, let $\mathbf{a} = \mathbf{v}|\mathbf{a}|$ be the *polar decomposition* of $\mathbf{a}$, having $|\mathbf{a}|$ as its positive part and $\mathbf{v}$ as its partial isometric part. Then, we denote by $L^2(\Gamma, \mathscr{B}, \tau)$ the Hilbert space completion of $\mathscr{B}$ in the topology given by the norm

$\| \cdot \| : \mathscr{B} \to [0, \infty)$,

$\mathbf{a} \mapsto \| \mathbf{a} \|_2 = [\tau(|\mathbf{a}|^2)]^{1/2}$.

We write $L^\infty(\Gamma, \mathscr{B}, \tau)$ for the W^*-algebra $\mathscr{B}$ and let $\| \cdot \|_\infty$ be the operator norm $\| \cdot \|$ of $\mathscr{B}$. We also write $L^1(\Gamma, \mathscr{B}, \tau)$ for the collection of all $\mathbf{a} \in L^0(\Gamma, \mathscr{B}, \tau)$ such that $\| \mathbf{a} \|_1 = \int_0^\infty \tau(e(d\lambda)\lambda) \equiv \tau(|\mathbf{a}|) < \infty$, where $\{e(\Lambda): \Lambda \text{ is Borel in } [0, \infty)\}$ is the resolution of the identity for $|\mathbf{a}|$. Notice that $L^1(\Gamma, \mathscr{B}, \tau)$ is not necessarily a Banach space since $\| \cdot \|_1$ is not necessarily a norm. However, $L^1(\Gamma, \mathscr{B}, \tau)$ is clearly a topological space.

We introduce next the notion of *conditioning* with respect to a W^*-subalgebra of $\mathscr{B}$.

2.11 *Definition*: Let $\mathscr{B}_0$ be a W^*-subalgebra of the W^*-algebra $\mathscr{B}$. Let $\mathbf{u}$ be an element of $L^1(\Gamma, \mathscr{B}, \tau)$. Then the *conditional* expectation of $\mathbf{u}$ given $\mathscr{B}_0$ is defined[11] as the operator $E(\mathbf{u}|\mathscr{B}_0)$ in $L^1(\Gamma, \mathscr{B}_0, \tau)$ satisfying

$$\tau(\mathbf{p}E(\mathbf{u}|\mathscr{B}_0)\mathbf{p}) = \tau(\mathbf{pup}), \quad \text{for all } \mathbf{p} \in \mathbf{p}(\mathscr{B}_0).$$

Notice that if τ is central then the last equation becomes

$$\tau(\mathbf{p}E(\mathbf{u}|\mathscr{B}_0)) = \tau(\mathbf{pu}), \quad \text{for all } \mathbf{p} \in p(\mathscr{B}_0).$$

The conditional expectation operator $E(\cdot|\mathscr{B}_0)$ is linear and it possesses the following additional properties:

(i) $E(\cdot|\mathscr{B}_0)$ is a contraction of $L^1(\Gamma, \mathscr{B}, \tau)$ onto $L^1(\Gamma, \mathscr{B}_0, \tau)$;

(ii) $E(\mathbf{u}^*|\mathscr{B}_0) = E(\mathbf{u}|\mathscr{B}_0)^*$, $\mathbf{u} \in L^1(\Gamma, \mathscr{B}, \tau)$;

(iii) $E(\cdot|\mathscr{B}_0)$ is positivity-preserving:

(iv) $E(1|\mathscr{B}_0) = 1$;

(v) $E(E(\mathbf{u}|\mathscr{B}_0)|\mathscr{B}_0) = E(\mathbf{u}|\mathscr{B}_0)$, $\mathbf{u} \in L^1(\Gamma, \mathscr{B}, \tau)$.

For a verification of the listed properties of $E(\cdot|\mathscr{B}_0)$, we refer to Ref. 11.

3. GENERALIZED STOCHASTIC FIELDS AND MARKOV PROPERTY

Let $(\Gamma, \mathscr{B}, \tau)$ and $L^0(\Gamma, \mathscr{B}, \tau)$ be as in the last section. An element of $L^0(\Gamma, \mathscr{B}, \tau)$ will be called an *operator-valued random variable*. Let $S(\mathscr{R}^d)$ be Schwartz space of C^∞ functions of rapid descent on $\mathscr{R}^d$. We shall say that ξ is an *operator-valued generalized random variable*, with underlying probability space, $(\Gamma, \mathscr{B}, \tau)$, provided that ξ is a continuous linear map from $S(\mathscr{R}^d)$ into the topological space $L^0(\Gamma, \mathscr{B}, \tau)$. By a *generalized stochastic field*, with underlying probability space $(\Gamma, \mathscr{B}, \tau)$ and indexed by $S(\mathscr{R}^d)$, we mean a noncommuting family

$$\mathscr{H}^0 = \{\mathbf{f} \mapsto \xi(\mathbf{f}): \mathbf{f} \in S(\mathscr{R}^d)\}$$

of operator-valued generalized random variables on $(\Gamma, \mathscr{B}, \tau)$. We denote by $\mathscr{H}^2$, the Hilbert space completion of $\mathscr{H}^0$ in the topology given by the norm:

$\| \cdot \|_{\mathscr{H}^2}: \mathscr{H}^0 \to [0, \infty)$

$\xi(\mathbf{f}) \mapsto \| \xi(\mathbf{f}) \|_{\mathscr{H}^2} = (\tau|\xi(\mathbf{f})|^2)^{1/2}$.

Here $\tau(|\xi(\mathbf{f})|^2) = \int_0^\infty \tau(e_{\mathbf{f}}(d\lambda))\lambda^2$, where $\{e_{\mathbf{f}}(\Lambda): \Lambda \text{ is Borel in } [0, \infty)\}$ is the resolution of the identity for $|\xi(f)|$. We remark that if τ is *central*, then $L^2(\Gamma, \mathscr{B}, \tau)$ and $\mathscr{H}^2$ are H^*-algebras of closed, densely defined operators on Γ. We shall write $\mathscr{H}^1$ for the collection of $\xi(f) \in \mathscr{H}^0$ such that $\tau(|\xi(\mathbf{f})|) < \infty$. Again $\mathscr{H}^1$ is not necessarily a Banach space,

The linear functional:

$m: \mathscr{H}^1 \to \mathbb{C}$,

$\mathbf{u} \mapsto m(\mathbf{u}) = \tau(\mathbf{u})$,

is called the *expectation functional*, and $m(\mathbf{u})$ is said to be the *expectation* value of $\mathbf{u} \in \mathscr{H}^1$.

The sesquilinear functional

$B: \mathscr{H}^2 \times \mathscr{H}^2 \to \mathbb{C}$,

$(\mathbf{u}, \mathbf{v}) \mapsto B(\mathbf{u}, \mathbf{v}) = \tau(\mathbf{u}^*\mathbf{v})$,

is called the *correlation functional*. Notice that each $\mathbf{u} \in \mathscr{H}^2$ has an expectation value because $\mathscr{H}^1 \supset \mathscr{H}^2$ [since $\tau(1) = 1$].

The Hilbert space $\mathscr{H}^2$ is isometrically isomorphic to the Hilbert space $H(\mathscr{R}^d)$ which is the completion of $S(\mathscr{R}^d)$ in the norm topology given by

$$\mathbf{f} \mapsto \|\mathbf{f}\|_{H(\mathscr{R}^d)} = \|\xi(\mathbf{f})\|_{\mathscr{H}^2}.$$

Next, let $V_{\mathbf{f}}|\xi(\mathbf{f})|$, $\mathbf{f} \in S(\mathscr{R}^d)$, be the polar decomposition of $\xi(\mathbf{f}) \in \mathscr{H}^0$, in which $V_{\mathbf{f}}$ is the partial isometric part and $|\xi(\mathbf{f})|$ is the positive part. Let

$$|\xi(\mathbf{f})| = \int_0^\infty e_{\mathbf{f}}(d\lambda)\,\lambda$$

be the spectral representation of $|\xi(\mathbf{f})|$. For D an open subset of $\mathscr{R}^d$, we denote by $\mathscr{B}(D)$ the W^*-subalgebra of $\mathscr{B}$ generated by the set: $\{V_{\mathbf{f}}, e_{\mathbf{g}}(\Lambda): \mathbf{f}, \mathbf{g} \in S(\mathscr{R}^d), \operatorname{supp}\mathbf{f} \subset D, \operatorname{supp}\mathbf{g} \subset D$ and Λ varies over all Borel subsets of $\mathscr{R}\}$. If C is an arbitray subset of $\mathscr{R}^d$, we set

$$\mathscr{B}(C) = \bigcap_{D \supset C} \mathscr{B}(D),$$

where the intersection is taken over all open subsets D of $\mathscr{R}^d$ each of which contains C. We now introduce the following notion of Markov property.

3.1 *Definition*: Let D be an open subset of $\mathscr{R}^d$ with complement D' and boundary ∂D. We shall say that the generalized stochastic field $\mathscr{H}^0$ is *Markovian* or has the *Markov property* if

$$E(\mathbf{u}|\mathscr{B}(D')) = E(\mathbf{u}|\mathscr{B}(\partial D))$$

for all $\mathbf{u} \in L^1(\Gamma, \mathscr{B}(D), \tau)$

3.2 *Remark*:

(i) If the generalized stochastic field $\mathscr{H}^0$ has the Markov property, we shall say that it is a *Markov field*.

(ii) For an example of a (noncommuting) Markov field of relevance to relativistic quantum field theory, we refer to Ref. 8.

(iii) For the rest of the paper, we shall make the following assumptions, for simplicity:

(a) the W^*-algebra $\mathscr{B}$ is generated by the partial isometric parts and the spectral projections of the positive parts of the canonical polar decomposition of the operators in $\mathscr{H}^0 = \{\xi(\mathbf{f}): \mathbf{f} \in S(\mathscr{R}^d)\}$;

(b) the net $\{\mathscr{B}(D): D$ is open in $\mathscr{R}^d\}$ is ordered under isotonous inclusion (i.e., if D_1, D_2 are open subsets of $\mathscr{R}^d$ such that $D_2 \supset D_1$, then $\mathscr{B}(D_1)$ is identifiable with a W^*-subalgebra of $\mathscr{B}(D_2)$ and the W^*-algebra $\mathscr{B}$ is the inductive limit[10] of the directed set $\{\mathscr{B}(D): D$ is open in $\mathscr{R}^d\}$ of W^*-subalgebras.

(c) if D is the union of open subsets $\{D_i\}$ of $\mathscr{R}^d$, then $(\cup \mathscr{B}(D_i))'' \supset \mathscr{B}(D)$.

(3.3) *Definition*: Let $\mathscr{H}^0 = \{\xi(\mathbf{f}): \mathbf{f} \in S(\mathscr{R}^d)\}$ be a Markov field. Then we shall say that $\mathscr{H}^0$ is a *regular* Markov field if for any open subset D of $\mathscr{R}^d$,

$$E(\mathscr{B}(D \cup D_0)|\mathscr{B}(D')) \subset \mathscr{B}(D_0),$$

where D_0 is an arbitrary closed subset of D', the complement of D, which contains ∂D.

3.4 *Remark*: We note here that generalized stochastic fields which satisfy assumptions (iii) (a), (b), and (c) of Remark 3.2 abound in constructive quantum field theory. Indeed, in the algebraic approach to quantum field theory, these same assumptions are customarily made.[18,19] Also, the Markov field discussed in Ref. 8 is an example of a regular Markov field.

4. MARKOV FIELDS AND MULTIPLICATIVE OPERATORS

In this section, we show how to generate new Markov fields from old ones. To this end, it is convenient to begin with the following notion.

4.1 *Definition*: We shall say that a measurable operator $\mathbf{a} \in \dot{L}^0(\Gamma, \mathscr{B}, \tau)$ is *multiplicative* provided that for every finite open covering $\{D_i\}_{i=1}^N$ of $\mathscr{R}^d$, there are strictly positive operators $\{\mathbf{a}_i\}_{i=1}$ with $\mathbf{a}_i \in L^0(\Gamma, \mathscr{B}(D_i), \tau)$, $i = 1,2,3,\ldots,N$ such that

$$\mathbf{a} = \mathbf{a}_1\mathbf{a}_2\cdots\mathbf{a}_N,$$

where $L^0(\Gamma, \mathscr{B}(D_i), \tau)$ is the collection of all linear closed, densely defined operators on Γ which are measurable with respect to $\mathscr{B}(D_i)$.

4.2 *Remark*: Next, for $\mathbf{c}$ any positive element of $L^\infty(\Gamma, \mathscr{B}, \tau)$, let $\tau_{\mathbf{c}}$ denote the functional on $\mathscr{B}$ given by

$$\tau_{\mathbf{c}}(\mathbf{u}) = \tau(\mathbf{uc}), \quad \text{for all } \mathbf{u} \in \mathscr{B}.$$

Then clearly $\tau_{\mathbf{c}}$ is a linear functional which is also positive if τ is central or if $\mathbf{c}$ belongs to Z. Furthermore $\tau_{\mathbf{c}}$ is bounded for

$$\begin{aligned}\|\tau_{\mathbf{c}}\| &= \sup_{0 \neq \mathbf{u} \in \mathscr{N}} \left(\frac{|\tau_{\mathbf{c}}(\mathbf{u})|}{\|\mathbf{u}\|}\right) \\ &= \sup_{0 \neq \mathbf{u} \in \mathscr{N}} \frac{|\tau(\mathbf{uc})|}{\|\mathbf{u}\|_\infty} \\ &\leqslant \sup_{0 \neq \mathbf{u} \in \mathscr{N}} \frac{\|\tau\|\|\mathbf{u}\|_\infty\|\mathbf{c}\|_\infty}{\|\mathbf{c}\|_\infty} \\ &= \|\tau\|\|\mathbf{c}\|_\infty.\end{aligned}$$

Hence convergence in measure with respect to τ implies convergence in measure with respect to $\tau_{\mathbf{c}}$.

4.3 *Theorem*: Let $\mathscr{H}^0 = \{\xi(\mathbf{f}): \mathbf{f} \in S(\mathscr{R}^d)\}$ be a regular Markov field over the probability space $(\Gamma, \mathscr{B}, \tau)$. Let $\mathbf{a} \in L^\infty(\Gamma, Z, \tau)$ be a right-invertible multiplicative measurable operator satisfying $\tau(\mathbf{a}) = 1$, and define $\tau_{\mathbf{a}}$ as in the foregoing. Then $\mathscr{H}^0 = \{\xi(\mathbf{f}): \mathbf{f} \in S(\mathscr{R}^d)\}$ is a Markov field over the probability space $(\Gamma, \mathscr{B}, \tau_{\mathbf{a}})$.

4.4 *Remark*:

(i) This theorem is a noncommutative extension of what happens in the theory of Markov stochastic processes where new Markov processes are generated from old ones by means of multiplicative functionals.[13]

(ii) We prove the theorem below after recalling a relevant result.

4.5 *Proposition*:

(i) Let $\mathscr{H}^0 = \{\xi(\mathbf{f}) : \mathbf{f} \in S(\mathscr{R}^d)\}$ be a *regular* Markov field indexed by $S(\mathscr{R}^d)$, let D be an open subset of $\mathscr{R}^d$, and let D_0 be a closed subset of the complement D' of D containing ∂D. Then

$$E(\mathscr{B}(D \cup D_0) | \mathscr{B}(D')) = \mathscr{B}(D_0).$$

(ii) Let D, D_0, and $\mathscr{H}^0 = \{\xi(\mathbf{f}) : \mathbf{f} \in S(\mathscr{R}^d)\}$ be as in (i), but with $\mathscr{H}^0$ not assumed necessarily a (regular) Markov field but satisfying (iii) (c) of Remark (3.2). Then

$$\mathscr{B}(D \cup D_0) = (\mathscr{B}(D) \cup \mathscr{B}(D_0))''$$

Proof:

(ii) Since $D_0 \supset \partial D$, the set $D \cup D_0$ is closed in $\mathscr{R}^d$. Let $\{D_n\}_{n=1}^{\infty}$ be a sequence of open subsets of $\mathscr{R}^d$ decreasing to D_0. Then $\{D \cup D_n\}_{n=1}^{\infty}$ is a sequence of open sets decreasing to $D \cup D_0$. By assumption (iii), it follows that

$$\mathscr{B}(D \cup D_n) = (\mathscr{B}(D) \cup \mathscr{B}(D_n))''$$

Next, recall that if $\mathscr{C}$ and $\{\mathscr{A}_n\}_{n=1}^{\infty}$ are W^*-subalgebras of $\mathscr{B}$, where $\{\mathscr{A}_n\}_{n=1}^{\infty}$ forms a decreasing sequence, i.e., $\mathscr{A}_n \supset \mathscr{A}_{n+1}$, $n = 1,2,\ldots,\infty$, then

$$\left(\bigcap_{n=1}^{\infty} (\mathscr{C} \cup \mathscr{A}_n)\right)'' = (\mathscr{C} \cup \bigcap_{n=1}^{\infty} \mathscr{A}_n)''.$$

Hence

$$\mathscr{B}(D \cup D_0) = \bigcap_{n=1}^{\infty} (\mathscr{B}(D) \cup \mathscr{B}(D_n))''$$

$$= (\mathscr{B}(D) \cup \bigcap_{n=1}^{\infty} \mathscr{B}(D_n))''$$

$$= (\mathscr{B}(D) \cup \mathscr{B}(D_0))''$$

Thus $\mathscr{B}(D \cup D_0)$ is generated by $\mathscr{B}(D) \cup \mathscr{B}(D_0)$. Finally, the proof of (i) follows trivially from (iii)(c) of Remark 3.2. This concludes the proof.

4.6 *Remark*: We wish next to give a proof of Theorem 4.3.

Proof of Theorem 4.3:

There are two probability spaces $(\Gamma, \mathscr{B}, \tau)$ and $(\Gamma, \mathscr{B}, \tau_{\mathbf{a}})$ for consideration. Denote by $E(\cdot|\mathscr{C})$ and $E_{\mathbf{a}}(\cdot|\mathscr{C})$ the conditional expectation operators with respect to the states τ and $\tau_{\mathbf{a}}$, respectively, where $\mathscr{C}$ is a W^*-subalgebra of $\mathscr{B}$.

Let D be an open subset of $\mathscr{R}^d$ with complement D' and boundary ∂D. We wish to prove that

$$E_{\mathbf{a}}(\mathbf{u}|\mathscr{B}(D')) = E_{\mathbf{a}}(\mathbf{u}|\mathscr{B}(\partial D))$$

for all $\mathbf{u} \in L^1(\Gamma, \mathscr{B}(D), \tau)$. To establish this, we need equivalently, to show that there is a uniquely determined $\tilde{\mathbf{u}} = E_{\mathbf{a}}(\mathbf{u}|\mathscr{B}(D'))$ in $L^0(\Gamma, \mathscr{B}(D')\tau)$, for every $\mathbf{u} \in L^1(\Gamma, \mathscr{B}(D), \tau)$, such that

$$\tau_{\mathbf{a}}(\tilde{\mathbf{u}}\mathbf{p}) = \tau_{\mathbf{a}}(\mathbf{p}\mathbf{u}\mathbf{p}), \quad \text{for all } \mathbf{p} \in p(\mathscr{B}(D')),$$

i.e.,

$$4.7\ \tau(\tilde{\mathbf{u}}\mathbf{p}\mathbf{a}) = \tau(\mathbf{p}\mathbf{u}\mathbf{p}\mathbf{a}), \quad \text{for all } \mathbf{p} \in p(\mathscr{B}(D')),$$

and, furthermore, that $\tilde{\mathbf{u}}$ does in fact belong to $\mathscr{B}(\partial D)$. Since the measurable multiplicative operator $\mathbf{a}$ belongs to Z, 4.7 is equivalent to

$$4.8\ \tau(\tilde{\mathbf{u}}\mathbf{a}\mathbf{p}) = \tau(\mathbf{u}\mathbf{a}\mathbf{p}), \quad \text{for all } \mathbf{p} \in p(\mathscr{B}(D')).$$

Let D_0 be an arbitrary open subset of $\mathscr{R}^d$ containing ∂D. Then the triple $\{D, D_0, D'^0\}$ is a finite open covering of $\mathscr{R}^d$, where D'^0 is the interior of D'. Then since $\mathbf{a}$ is a multiplicative measurable operator, we have that there are positive measurable operators $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ with $\mathbf{a}_1 \in L^0(\Gamma, \mathscr{B}(D), \tau)$, $\mathbf{a}_2 \in L^0(\Gamma, \mathscr{B}(D), \tau)$, and $\mathbf{a}_3 \in L^0(\Gamma, \mathscr{B}(D'^0), \tau)$ such that

$$4.9\ \mathbf{a} = \mathbf{a}_1\mathbf{a}_2\mathbf{a}_3.$$

Employing 4.9 in 4.8, we have

$$\tau(\mathbf{p}\tilde{\mathbf{u}}\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3\mathbf{p}) = \tau(\mathbf{p}\mathbf{u}\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3\mathbf{p}) \quad \text{for all } \mathbf{p} \in p(\mathscr{B}(D'))$$

or

$$\tau(\mathbf{p}E(\tilde{\mathbf{u}}\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3|\mathscr{B}(D'))\mathbf{p}) = \tau(\mathbf{p}E(\mathbf{u}\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3|\mathscr{B}(D'))\mathbf{p}), \quad \text{for all } \mathbf{p} \in p(\mathscr{B}(D')).$$

Hence the sought-for $\tilde{\mathbf{u}}$ must satisfy

$$E(\tilde{\mathbf{u}}\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3|\mathscr{B}(D')) = E(\mathbf{u}\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3|\mathscr{B}(D')).$$

Using the properties of the conditional expectation operator, we have, since $\mathbf{a}_3 \in L^0(\Gamma, \mathscr{B}(D'), \tau)$ and $\tilde{\mathbf{u}}$ also belongs to $L^0(\Gamma, \mathscr{B}(D'), \tau)$ when it exists, that

$$\tilde{\mathbf{u}}E(\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))\mathbf{a}_3 = E(\mathbf{u}\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))\mathbf{a}_3.$$

Since $\mathbf{a}$ is right-invertible, so is $\mathbf{a}_3$. Hence

$$4.10\ \tilde{\mathbf{u}}E(\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D')) = E(\mathbf{u}\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D')).$$

But the right-inverse of $E(\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))$ is a densely defined measurable operator in $L^0(\Gamma, \mathscr{B}(D'), \tau)$. Hence, we have

$$4.11\quad \tilde{\mathbf{u}} = E(\mathbf{u}\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))[E(\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))]^{-1},$$ where the inverse here is the right-inverse of the indicated operator.

Set $\bar{D}_0 \cap D' = D^c$, where $\bar{D}_0$ is the closure of D_0.

Since $\mathscr{H}^0 = \{\xi(\mathbf{f}) : \mathbf{f} \in S(\mathscr{R}^d\}$ is assumed to be a regular Markov field on the probability space $(\Gamma, \mathscr{B}, \tau)$ part (i) of Proposition 4.5 is available. Applying the latter result with D and D^c replacing the subsets of $\mathscr{R}^d$ occurring there, it follows that $[E(\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))]^{-1}$ and $E(\mathbf{u}\mathbf{a}_1\mathbf{a}_2|\mathscr{B}(D'))$ are in $L^0(\Gamma, \mathscr{A}(D^c), \tau)$ which is contained in $L^0(\Gamma, \mathscr{B}(D_0), \tau)$. Since D_0 is an arbitrary open subset of $\mathscr{R}^d$ containing ∂D we conclude that $\tilde{\mathbf{u}}$ is in $L^0(\Gamma, \mathscr{B}(\partial D), \tau)$. Thus defining $\tilde{\mathbf{u}}$ by 4.10 we are assured of its existence, together with the desired measurability properties; furthermore, $\tilde{\mathbf{u}}$ defined in this way satisfies 4.7. The uniqueness of $\tilde{\mathbf{u}}$ is trivial. This concludes the proof.

4.12 *Remark*: (i) The proof shows that, in fact,

$$E_a(\mathbf{u}|\mathscr{B}(D')) = E(\mathbf{u}\mathbf{a}|\mathscr{B}(D'))(E(\mathbf{a}|\mathscr{B}(D'))]^{-1}$$

for any right-invertible positive multiplicative measurable operator g in $L^\infty(\Gamma, Z, \tau)$ such that $\tau(\mathbf{a}) = 1$, as is clear from 4.10 by right-multiplication by $\mathbf{a}_3$, which belongs to $L^0(\Gamma, \mathscr{B}(D'), \tau)$.

(ii) It is hoped that Theorem 4.3 would be useful in the Euclidean approach to interacting fermion quantum field theory.

[1]I.E. Segal, "A Noncommutative Extension of Abstract Integration," Ann. Math. N.Y. **57**, 401–57 (1953).
[2]I.E. Segal, Correction to "A Noncommutative Extension of Abstract Integration," Ann. Math. N.Y. **58**, 595–6 (1953).
[3]I.E. Segal, "Algebraic Integration Theory," Bull. Am. Math. Soc. **71**, 419–89 (1965).
[4]R.A. Kunze, "Fourier Transforms on Locally-Compact Unimodular Groups, Trans. Am. Math. Soc. **89**, 519–40 (1958).
[5]W.F. Stinespring, "Integration Theorems for Gauges and Duality for Unimodular Groups," Trans. Am. Math. Soc. **90**, 15–56 (1959).
[6]I.E. Segal, "Tensor Algebras over Hilbert Spaces II," Ann. Math. **63**, 419–89 (1965).
[7]L. Gross,"Existence and Uniqueness of Physical Ground States," J. Func. Anal. **10**, 52–109 (1972).
[8]I.F. Wilde, "The Free Fermion Field as a Markov Field," J. Func. Anal. **15**, 12–21 (1974).
[9]E. Nelson, "Notes on Noncommutative Integration," J. Funct. Anal. **15**, 103–16 (1974).
[10]H.A. Dye, "The Radon–Nikodym theorem for Finite Rings of Operators," Trans. Am. Math. Soc. **72**, 243–80 (1952).
[11]S. Sakai, *C*-Algebras and W*-Algebras* (Springer-Verlag, Berlin, 1971).
[12]S. Gudder, and J.-P. Marchand, "Noncommutative Probability on von Neumann Algebras," J. Math. Phys. **13**, 799–806 (1972).
[13]I.I. Gihman and A.V. Skorohod, *The Theory of Stochastic Processes* II (Springer-Verlag, Berlin, 1975).
[14]G. Velo, and A. Wightman (Editors) *Constructive Quantum Field Theory, Lectures Notes in Physics* (Springer-Verlag, Berlin, 1973).
[15]B. Simon: *The $P(\Phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, New Jersey, 1974).
[16]F. Guerra, L. Rosen, and B. Simon "Nelson's Symmetry and the Infinite Volume Behaviour of the Vacuum in $P(\Phi)_2$," Commun. Math. Phys. **27**, 10–22 (1972).
[17]M.J. Maczynski, "Noncommutative probability on von Neumann Algebras," *Lecture Notes in Mathematics* Vol. **472**, 107–117 (1975).
[18]R. Haag and D. Kastler "An Algebraic Approach to Quantum field Theory," J. Math. Phys. **5**, 848–61 (1964).
[19]R. Haag, "Quantum Field Theory" in *Mathematics of Contemporary Physics*, edited by R.F. Streater (Academic, London and New York, 1972).

A note on the Lorentz transformation

W.-H. Steeb
Universität Paderborn, FB 6, Theoretische Physik, 4790 Paderborn, West Germany

(Received 28 August 1978; revised manuscript received 27 November 1978)

Using Lie theory of one-parameter transformation group, we show that the (linear) Lorentz transformation can be embedded into a class of nonlinear transformations.

1. INTRODUCTION

Recently, in a series of papers,[1-5] the author has investigated a connection between Lie's theory of one-parameter groups and autonomous nonlinear systems of differential equations. For a given system of linear differential equations $\dot{x} = \xi(x) = Ax$ (where A is an $n \times n$ matrix with constant coefficients), the nonlinear system $\dot{x} = \eta(x)$ is constructed via the relation $[X, Y] = f(x)Y$. $[X, Y]$ denotes the commutator (Lie bracket) of the C^∞-vector fields (what in physics are called infinitesimal generators) X and Y. $f(x)$ is a smooth function. The vector fields X and Y are written in local coordinates as $X = \xi_1 \partial/\partial x_1 + \cdots + \xi_n \partial/\partial x_n$ and $Y = \eta_1 \partial/\partial x_1 + \cdots + \eta_n \partial/\partial x_n$. In the cited papers the described approach has been applied to systems containing periodic orbits and limit cycles. For example, the equation of motion of the harmonic oscillator $\dot{x}_1 = \dot{x}_2; \dot{x}_2 = -x_1$ is associated with the vector field $X = x_2 \partial/\partial x_1 - x_1 \partial/\partial x_2$. The vector field Y obtained via the equation $[X, Y] = f(x)Y$ leads to nonlinear systems $\dot{x} = Y(x)$ which contain limit cycles. The limit cycles themselves are described by the equation of motion for the harmonic oscillator.

In the present paper we apply the described approach to an autonomous system of differential equations which is important in the theory of relativity. This means, we study the system of linear differential equations ($\dot{x}_1 = x_2$, $\dot{x}_2 = x_1$), the associated vector field $X = x_2 \partial/\partial x_1 + x_1 \partial/\partial x_2$ and construct the most general vector field Y which commutes with X. Then the physical meaning of the vector field Y (and its associated nonlinear system of differential equations) is given.

2. MATHEMATICAL PRELIMINARIES

First of all, let us briefly recall some well-known results. Consider the system

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = x_1, \tag{2.1}$$

with $\dot{x}_i = dx_i/d\lambda$. Let $M = \mathbb{R}^2$ (M = manifold). The system of differential equations (2.1) induces the vector field $X = x_2 \partial/\partial x_1 + x_1 \partial/\partial x_2$. The vector field X is complete on $M = \mathbb{R}^2$. Moreover, the vector field X has the following properties. Let $X = x_2 \partial/\partial x_1 + x_1 \partial/\partial x_2$ and let $L_X(\cdot)$ denote the Lie derivative of a geometric object (i.e., function, vector field, tensor field) with respect to X. Then

$$L_X(x_1^2 - x_2^2) = X(x_1^2 - x_2^2) = 0, \tag{2.2a}$$

$$L_X\left(x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2}\right) = \left[X, x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2}\right] = 0, \tag{2.2b}$$

$$L_X(dx_1 \wedge dx_2) = 0, \tag{2.2c}$$

$$L_X(dx_1 \otimes dx_1 - dx_2 \otimes dx_2) = 0. \tag{2.2d}$$

When we apply the usual rules of the Lie derivative (compare [6] and references herein), the relations (a) through (d) can easily be obtained. Equation (2.2a) tells us that the quantity $x_1^2 - x_2^2$ is the (global) first integral of X. From Eq. (2.2a) it follows that $Xf(x_1^2 - x_2^2) = 0$, where f is a smooth function. The vector fields $x_2 \partial/\partial x_1 + x_1 \partial/\partial x_2$ and $x_1 \partial/\partial x_1 + x_2 \partial/\partial x_2$ generate one-parameter groups Φ_λ and Ψ_μ, respectively. Because of Eq. (2.2b), it follows that $\Phi_\lambda \circ \Psi_\mu = \Psi_\mu \circ \Phi_\lambda$. Equation (2.2c) tells us that the flow Φ_λ (one-parameter group) generated by X is an area-preserving mapping. Finally, the vector field X is a Killing vector field with respect to the (2,0) tensor field $g = dx_1 \otimes dx_1 - dx_2 \otimes dx_2$ [Eq. (2.2d)]. Note that the vector fields $\partial/\partial x_1$ and $\partial/\partial x_2$ are Killing vector fields (with respect to g), too. The Killing vector fields lead to conserved quantities. Moreover, the vector fields $\{\partial/\partial x_1, \partial/\partial x_2, x_2 \partial/\partial x_1 + x_2 \partial/\partial x_1\}$ form a Lie algebra as it must be. The solution of the system of differential equations (2.1) is given by $x_1 = x_{10} \cosh\lambda + x_{20} \sinh\lambda$, $x_2 = x_{20} \cosh\lambda + x_{10} \sinh\lambda$ [$x_{10} = x_1(\lambda = 0)$, $x_{20} = x_2(\lambda = 0)$]. In other words, we have the one-parameter transformation group $\Phi: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2: x_1 \to x_1 \cosh\lambda + x_2 \sinh\lambda$, $x_2 \to x_2 \cosh\lambda + x_1 \sinh\lambda$.

The most general vector field Y which commutes with X (i.e., $[X, Y] = 0$) has the form (in local coordinates)

$$Y = [x_1 f_2(x_1^2 - x_2^2) + x_2 f_1(x_1^2 - x_2^2)] \frac{\partial}{\partial x_1} + [x_2 f_2(x_1^2 - x_2^2) + x_1 f_1(x_1^2 - x_2^2)] \frac{\partial}{\partial x_2}, \tag{2.3}$$

where both f_1 and f_2 are smooth functions. The method for obtaining the vector field Y is well known.[7-10] Let us describe a rather different method for obtaining the vector field Y. To obtain the vector field Y we consider the Abelian Lie algebra $\{X, I = x_1 \partial/\partial x_1 + x_2 \partial/\partial x_2\}$ and the rule $[A, fB] = (Af)B + f[A, B]$. Since $Xf(x_1^2 - x_2^2) = 0$, we find that

$$[X, f_1(x_1^2 - x_2^2)X + f_2(x_1^2 - x_2^2) I] = 0. \tag{2.4}$$

Let $M = \mathbb{R}^2$. The vector field X generates a one-parameter group of transformations. On the other hand, in general, the vector field Y is not complete. Obviously, whether or not the vector field Y is complete depends on the form of both f_1 and f_2. The condition $[X, Y] = 0$ is an integrability condition. The geometrical meaning is that the flow Φ_λ (associated with X) and the (local) flow Γ_μ (associated with Y) commute ($\Phi_\lambda \circ \Gamma_\mu = \Gamma_\mu \circ \Phi_\lambda$) is an appropriate domain.

 0022-2488/79/081684-03$01.00

The vector field Y is associated with the autonomous system of differential equations

$$\dot{x}_1 = x_1 f_2(x_1^2 - x_2^2) + x_2 f_1(x_1^2 - x_2^2),$$
$$\dot{x}_2 = x_2 f_2(x_1^2 - x_2^2) + x_1 f_1(x_1^2 - x_2^2). \tag{2.5}$$

Let us investigate a particular case, namely $f_1(x_1,x_2) = 1$ and $f_2(x_1,x_2) = x_1^2 - x_2^2 + k$, where $k \in \mathbb{R}$. Now we discuss the nonlinear system ($\dot{x}_i = dx_i/d\mu$)

$$\dot{x}_1 = x_2 + x_1(x_1^2 - x_2^2 + k), \quad \dot{x}_2 = x_1 + x_2(x_1^2 - x_2^2 + k). \tag{2.6}$$

The associated vector field

$$Y = \left(x_2 + x_1(x_1^2 - x_2^2 + k)\right)\frac{\partial}{\partial x_1} + \left(x_1 + x_2(x_1^2 - x_2^2 + k)\right)\frac{\partial}{\partial x_2} \tag{2.7}$$

has the properties:

$$L_Y(x_1^2 - x_2^2 + k) = 2(x_1^2 - x_2^2)(x_1^2 - x_2^2 + k), \tag{2.8a}$$

$$L_Y(dx_1 \wedge dx_2) = 2((x_1^2 - x_2^2 + k) + x_1^2 + x_2^2))\, dx_1 \wedge dx_2, \tag{2.8b}$$

$$L_Y g = 2(x_1^2 - x_2^2 + k)\, g + 4x_1^2\, dx_1 \otimes dx_1 + 4x_2^2\, dx_2 \otimes dx_2 - 4x_1x_2 \times (dx_1 \otimes dx_2 + dx_2 \otimes dx_1). \tag{2.8c}$$

If we consider the equation $x_1^2 - x_2^2 + k = 0$ (hyperbola), then the right-hand side of Eq. (2.8a) vanishes.

The equation $x_1^2 - x_2^2 = -k$ ($k \neq 0$) defines the hyperbola. If $k = 0$, then we obtain the straight lines $x_1 = \pm x_2$ (the asymptotes). The hyperbola can be viewed as a one-dimensional integral manifold of the system given above [Eq. (2.6)]. Using the ansatz $r^2 = x_1^2 - x_2^2$, we get the differential equation

$$\frac{dr^2}{d\mu} = r^2(r^2 + k). \tag{2.9}$$

This equation can easily be solved. Let $k = 0$. Then the μ independent solution (i.e., $dr^2/d\mu = 0$)is given by $(r^2)^2 = 0$, i.e., $x_1^2 - x_2^2 = 0$. The μ dependent solution (local flow) is given by $r^2 = r_0^2/(1 - 2r_0^2\mu)$. As $\mu \to \infty$ (or $\mu \to -\infty$) we have $r^2 = 0$ or $x_1^2 - x_2^2 = 0$. Now let $k \neq 0$.Then we find the μ independent solution $r^2(r^2 + k) = 0$ or $(x_1^2 - x_2^2) \times (x_1^2 - x_2^2 + k) = 0$. The μ dependent solution (local flow) takes the form

$$r^2 = \frac{r_0^2\, k \exp(2k\mu)}{r_0^2(1 - \exp(2k\mu)) + k}. \tag{2.10}$$

If $r_0^2 = 0$, then $r^2 = 0$. If $r_0^2 = -k$, then $r^2 = k$. Now let $r_0^2 \neq 0$ and $r_0^2 \neq -k$. Let $k > 0$. As $\mu \to \infty$, we find $r^2 = -k$. As $\mu \to -\infty$, we find $r^2 = 0$. Now let $k < 0$. As $\mu \to \infty$, we find $r^2 = 0$. As $\mu \to -\infty$, we find $r^2 = -k$.

To sum up, as $\mu \to \infty$ (or $\mu \to -\infty$) the solution of the nonlinear equation either tends to the hyperbola ($x_1^2 - x_2^2 = -k$) or to the straight lines $x_1 = \pm x_2$ according to the value of k and to the initial values x_{10} and x_{20}.

To complete the mathematics, we put $\tanh\alpha = x_2/x_1$ and we find that $\dot{\alpha} = 1$. It follows that $\alpha = \alpha_0 + \mu$. Note that $x \to \operatorname{arctanh} x$ is a multiple valued function.

3. PHYSICAL SIGNIFICANCE

Let us now discuss the physical significance (obviously in a speculative manner) in the theory of special relativity of the described approach. We make the assumption that the reference frame Σ' is moving with a uniform velocity v along the X axis of the reference frame Σ. A 4-vector A is defined as any system possessing components $A_0\, A_1, A_2, A_3$ which behave like the space–time coordinates ct, x, y, z under Lorentz transformation:

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix}\begin{pmatrix} ct \\ x \end{pmatrix}, \tag{3.1}$$

where $y' = y$, $z' = z$, $\gamma = 1/(1 - \beta^2)^{1/2}$ and $\beta = v/c$ (c = speed of light). Examples of 4-vectors are the 4-velocity and the 4-momentum. The 4-velocity can be written as $(c\gamma, \mathbf{u}\gamma)$, where $\mathbf{u}$ is the usual three-dimensional velocity vector. If we multiply the 4-velocity by $m_0 c$ (m_0 rest mass), we obtain the momentum 4-vector $(mc^2, m\mathbf{u}c) = (E, p_x c, p_y c, p_z c)$ with $m = m_0\gamma$ and $p_x = mu_x$. E is the total energy. The invariant scalar product of two 4-vectors, say A and B, is, in Minkowski coordinates, $AB = -A_0B_0 + A_1B_1 + A_2B_2 + A_3B_3$. The scalar product of the 4-velocity with itself is given by $-c^2$ (timelike). Consequently, the scalar product of the 4-momentum with itself is given by $-m_0^2c^4$. It follows that

$$E^2 - \mathbf{p}^2c^2 = m_0^2c^4, \tag{3.2}$$

where $\mathbf{p} = (p_x, p_y, p_z)$. A photon has zero rest mass ($m_0 = 0$), and therefore $E^2 - \mathbf{p}^2c^2 = 0$ in this case. $E^2 - \mathbf{p}^2c^2$ remains invariant under Lorentz transformations. In a closed system, with no energy or matter entering or leaving, Eq. (4.2) connects the principle of conservation of energy with that of conservation of momentum. m_0 and c are constants. In other words, the momentum 4-vector of a closed system does not change in length or direction, no matter what happens inside the system.

Consequently, we identify the quantities x_1, x_2, and λ as follows: We put $x_1 = A_1$, $x_2 = A_0 (A'_2 = A_2, A'_3 = A_3)$, $\cosh\lambda = \gamma$, and $\sinh\lambda = -\beta\gamma$. It follows that $\lambda = \operatorname{arcsinh}(-\beta\gamma)$. If we consider the 4-momentum, then we additionally put $k = m_0^2c^4$. Let us consider the nonlinear system [Eq. (2.6)]. Let $x_1 = x$ and $x_2 = ct$. Instead of $x^2 - c^2t^2 = x_0^2 - ct_0^2$, we have

$$x^2 - c^2t^2 = \frac{(x_0^2 - c^2t_0^2)k \exp(2k\mu)}{(x_0^2 - c^2t_0^2)(1 - \exp(2k\mu)) + k}. \tag{3.3}$$

The properties of this solution have been widely discussed in Sec. 2. Now let $x_1 = p_x c$ and $x_2 = E$. In the theory of special relativity we have $E^2 - p^2c^2 = m_0^2c^4$. The nonlinear system describes a "dissipative system" (damping terms), i.e., the divergence of the vector field Y does not vanish. We obtain $\operatorname{div} Y = 2(x_1^2 - x_2^2 + k) + 2(x_1^2 - x_2^2)$. However, as $\mu \to \infty$, the system is no longer dissipative and it is described by $\dot{x}_1 = x_2$, $\dot{x}_2 = x_1$. Hence we describe an open system with energy or matter entering or leaving the system. Whether energy (or matter) enters or leaves the system depends on the form of the functions f_1 and f_2. Note that the scalar product of Y is given by $E^2 - p_x^2c^2 + (E^2 - p_x^2c^2)(E^2 - p_x^2c^2 - m_0^2c^4)$.

[1]W.-H. Steeb, Phys. Lett. A **62**, 211 (1977).
[2]W.-H. Steeb, Lett. Math. Phys. **2**, 171 (1977).
[3]W.-H. Steeb, J. Phys. A **10**, L221 (1977).
[4]W.-H. Steeb, Int. J. Theor. Phys. **16**, 671 (1977).
[5]W.-H. Steeb, Physica A **95**, 181 (1978).
[6]W.-H. Steeb, Z. Naturforschung A **33**, 742 (1978).
[7]A. Cohen, *An Introduction to the Lie Theory of One-Paramter Groups* (Heath, Boston, 1911).
[8]E.L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[9]L.P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).
[10]G.W. Bluman and J.D. Cole, *Similarity Methods for Differential Equations* (Springer-Verlag, New York, 1974).

The inverse problem for random sources[a)]

A. J. Devaney
Schlumberger–Doll Research Center, P. O. Box 307, Ridgefield, Connecticut 06877

(Received 23 January 1979)

The problem of deducing the statistical structure of a localized random source $\rho(\mathbf{r})$ of the reduced wave equation from measurements of the field external to the source is addressed for the case when the measurements yield the autocorrelation function of the field at all pairs of points exterior to the source volume and the quantity to be determined is the source's autocorrelation function $R_\rho(\mathbf{r}_1,\mathbf{r}_2) = \langle \rho^*(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle$. This problem is shown to be equivalent to that of determining R_ρ from the autocorrelation function of the field's radiation pattern and is found, in general, not to admit a unique solution due to the possible existence of nonradiating sources within the source volume. Notable exceptions are the class of delta correlated (incoherent) sources whose intensity profiles are shown to be uniquely determined from the data and the class of quasihomogeneous sources whose coherence properties can be determined if their intensity profiles are known and vice versa.

1. INTRODUCTION

An inverse problem of interest in optics and acoustics is that of deducing a deterministic source $\rho(\mathbf{r})$ of the reduced wave equation

$$(\nabla^2 + k_0^2)\psi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{1.1}$$

from measurements of the field ψ at points external to the region of localization V of the source. In the most favorable case ψ will be exactly known everywhere outside V. In this case the inverse problem reduces to that of determining $\rho(\mathbf{r})$ from the value of its Fourier transform

$$\tilde{\rho}(\mathbf{k}) = \int_V d^3r\, \rho(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{1.2}$$

evaluated in $\mathbf{k}$ space on the surface of a sphere centered at the origin ($\mathbf{k} = \mathbf{0}$) and of radius k_0. This conclusion follows from the fact that $\psi(\mathbf{r})$ is uniquely determined everywhere outside V by its *radiation pattern*,[1] which, in turn, is equal[2] to the above stated boundary value of the sources transform; i.e.,

$$\psi(\mathbf{r}) \sim \tilde{\rho}(k_0\hat{\mathbf{r}}) e^{ik_0 r}/r, \quad (k_0 r \to \infty), \tag{1.3}$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$.

The inverse problem described above does not admit a unique solution due to the possible existence of so-called *nonradiating sources*[3] within the source volume. Such sources possess Fourier transforms which vanish identically when $|\mathbf{k}| = k_0$ and thus produce fields which vanish everywhere outside their region of localization. It follows that solutions to the inverse source problem can be determined only up to an unknown additive nonradiating part which must be specified by information other than field data.

In many applications the source $\rho(\mathbf{r})$ and, hence, the field $\psi(\mathbf{r})$ will not be deterministic (i.e., perfectly coherent) but rather will be realizations of random processes that are characterized by the source autocorrelation function[4]

$$R_\rho(\mathbf{r}_1,\mathbf{r}_2) = \langle \rho^*(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle \tag{1.4}$$

and higher order moments. Within this context of "random sources" the inverse problem becomes that of deducing the *statistical structure* of the source from physically realizable measurements of the radiated field.

In this paper we address the inverse problem for random sources for the case when the measurements yield the autocorrelation function[5] of the field

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2) = \langle \psi^*(\mathbf{r}_1)\psi(\mathbf{r}_2)\rangle \tag{1.5}$$

at all pairs of points exterior to the source volume and the quantity to be determined is the source's autocorrelation function $R_\rho(\mathbf{r}_1,\mathbf{r}_2)$. In Sec. 2 it is shown that the autocorrelation function of the field is uniquely determined everywhere outside the source region by the autocorrelation function of the radiation pattern and vice-versa. This latter quantity is shown to be equal to the (six-dimensional) Fourier spectrum of the source autocorrelation function[6]

$$\Phi(\mathbf{k}_1,\mathbf{k}_2) = \langle \tilde{\rho}^*(\mathbf{k}_1)\tilde{\rho}(\mathbf{k}_2)\rangle \tag{1.6}$$

evaluated on the four-dimensional surface $|\mathbf{k}_1| = |\mathbf{k}_2| = k_0$. The inverse problem for random sources is thus found to reduce to that of determining $\Phi(\mathbf{k}_1,\mathbf{k}_2)$ [and, hence, $R_\rho(\mathbf{r}_1,\mathbf{r}_2)$] from its boundary value on the surface $|\mathbf{k}_1| = |\mathbf{k}_2| = k_0$.

It is shown in Sec. 3 that it is not possible in general to uniquely determine the spectrum $\phi(\mathbf{k}_1,\mathbf{k}_2)$ from its boundary value as determined by the autocorrelation function of the radiation pattern. Thus, like its deterministic counterpart, the inverse problem for random sources does not in general admit a unique solution. A notable exception is found to be the class of delta correlated sources for which the intensity profile $\langle |\rho(\mathbf{r})|^2\rangle$ can be uniquely determined from the known boundary value of $\Phi(\mathbf{k}_1,\mathbf{k}_2)$.

Finally, in Sec. 4, the inverse problem is formulated for so-called quasihomogeneous sources.[7] These sources are locally statistically homogeneous and are characterized by an autocorrelation function which factors into the product of the intensity profile of the source with a normalized autocor-

[a)] Preliminary results of this investigation were presented at the 1978 annual meeting of the Optical Society of America. [Abstract Th74, J. Opt. Soc. Am. **68**, 1421 (1978)]

 0022-2488/79/081687-05$01.00

relation function $g(\mathbf{r}_2 - \mathbf{r}_1)$ of a strictly statistically homogeneous source. It is shown for such sources that if either the intensity profile or the normalized autocorrelation function $g(\mathbf{r}_2 - \mathbf{r}_1)$ is known, the other can be uniquely determined from the autocorrelation function of the radiation pattern.

2. FORMULATION OF THE INVERSE PROBLEM

Throughout this paper we shall restrict our attention to that class of random sources whose realizations $\rho(\mathbf{r})$ are piecewise continuous and localized within a volume V which, for convenience, we take to be a sphere of radius R_0 centered at the origin. The field generated by any such realization is identified with that particular solution of Eq. (1.1) which behaves as an outgoing spherical wave at infinity, i.e., such that

$$\psi(\mathbf{r}) \sim \hat{\psi}(\hat{\mathbf{r}}) e^{ik_0 r}/r, \quad (k_0 r \to \infty), \tag{2.1}$$

where $\hat{\psi}(\hat{\mathbf{r}})$ is the radiation pattern of the field evaluated in the direction $\hat{\mathbf{r}}$ of the field point $\mathbf{r} = r\hat{\mathbf{r}}$. The appropriate solution of Eq. (1.1) satisfying the asymptotic condition (2.1) is

$$\psi(\mathbf{r}) = \int_V d^3r' \, \rho(\mathbf{r}') e^{ik_0|\mathbf{r} - \mathbf{r}'|}/|\mathbf{r} - \mathbf{r}'|. \tag{2.2}$$

It is easily verified that the field given in Eq. (2.2) satisfies the asymptotic condition (2.1) with the radiation pattern given by

$$\hat{\psi}(\hat{\mathbf{r}}) = \int_V d^3r' \, \rho(\mathbf{r}') e^{-ik_0 \hat{\mathbf{r}} \cdot \mathbf{r}'} = \tilde{\rho}(k_0 \hat{\mathbf{r}}). \tag{2.3}$$

As mentioned in the Introduction the radiation pattern $\hat{\psi}(\hat{\mathbf{r}})$ not only yields information about the far field but, in fact, uniquely specifies $\psi(\mathbf{r})$ everywhere outside the source volume and vice versa.[2] This one-to-one correspondence between the radiation pattern and the value of the field outside V is extremely important for the inverse problem and, thus, will now be established with the aid of the well-known multipole expansion of the field. This expansion, which can be obtained by expanding the Green function appearing in Eq. (2.2) into a series of spherical wave eigenfunctions of the reduced wave equation, converges everywhere outside V (i.e., for $r > R_0$) and is given by[8]

$$\psi(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_n^m h_n(k_0 r) Y_n^m(\theta,\phi). \tag{2.4}$$

Here $h_n(k_0 r)$ is the spherical Hankel function of the first kind of order n, $Y_n^m(\theta,\phi)$ is the spherical harmonic of degree n and order m, and (r,θ,ϕ) are the spherical polar coordinates of the field point $\mathbf{r}$. The expansion coefficients (multipole moments) a_n^m can be determined from the value of the field given over the surface of a sphere of radius $R > R_0$ by means of the formula

$$a_n^m = \frac{1}{h_n(k_0 R)} \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin\theta \, \psi(R\hat{\mathbf{r}}) Y_n^{m*}(\theta,\phi). \tag{2.5}$$

It follows from Eqs. (2.4) and (2.5) that there is a one-to-one correspondence between the set of multipole moments $\{a_n^m\}$ $(n = 0, 1,\ldots; m = -n, -n+1,\ldots,n)$ and the value of the field specified at all points lying outside the source volume. Moreover, by asymptotically expanding both sides of Eq. (2.4) we find that

$$\hat{\psi}(\hat{\mathbf{r}}) = \frac{1}{k_0} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_n^m Y_n^m(\theta,\phi), \tag{2.6}$$

from which we conclude that there is also a one-to-one correspondence between the set of multipole moments and the radiation pattern and, hence, between the radiation pattern and the value of the field at all points lying outside V.

The deterministic inverse source problem is that of deducing the source $\rho(\mathbf{r}')$ from the field $\psi(\mathbf{r})$ specified everywhere outside the source volume. This problem thus consists mathematically of solving the integral equation (2.2) for $\rho(\mathbf{r}')$ in terms of the value of $\psi(\mathbf{r})$ specified everywhere outside V. However, the one-to-one correspondence between the value of the field outside V and the radiation pattern shows that this problem is equivalent to that of determining the source transform $\tilde{\rho}(\mathbf{k})$ for all values of the wave vector $\mathbf{k}$ from its boundary value $\tilde{\rho}(k_0 \hat{\mathbf{r}})$ as given by the radiation pattern via Eq. (2.3).

When the source $\rho(\mathbf{r})$ is a random process the inverse problem becomes that of deducing the statistical structure of this process from physically realizable field measurements performed exterior to the source region. In practice, especially at optical frequencies, the field measurements will consist of interference experiments from which the autocorrelation function $\Gamma(\mathbf{r}_1,\mathbf{r}_2)$ [cf. Eq. (1.5)] can be determined. In the ideal case, which we address here, $\Gamma(\mathbf{r}_1,\mathbf{r}_2)$ will be known for all pairs of points lying outside the source volume. It is natural then to define the inverse problem for random sources to be that of deducing the source autocorrelation from this information. Mathematically, this consists of solving the integral equation

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2) = \int_V d^3r_1' \int_V d^3r_2' \, R_\rho(\mathbf{r}_1',\mathbf{r}_2') \times \frac{e^{ik_0|\mathbf{r}_2 - \mathbf{r}_2'|}}{|\mathbf{r}_2 - \mathbf{r}_2'|} \cdot \frac{e^{-ik_0|\mathbf{r}_1 - \mathbf{r}_1'|}}{|\mathbf{r}_1 - \mathbf{r}_1'|} \tag{2.7}$$

for $R_\rho(\mathbf{r}_1',\mathbf{r}_2')$ in terms of $\Gamma(\mathbf{r}_1,\mathbf{r}_2)$ given for all pairs of points outside V, and thus is completely analogous to the deterministic problem of inverting Eq. (2.2) for $\rho(\mathbf{r}')$given $\psi(\mathbf{r})$ everywhere outside V.

It follows from Eqs. (2.4) and (2.5) that the autocorrelation function of the field at all pairs of points external to the source region is uniquely specified by the various statistical moments $\langle a_n^{m*} a_{n'}^{m'} \rangle$ and vice versa. In virtue of Eq. (2.6) a similar one-to-one correspondence exists between these statistical moments and the autocorrelation function $\langle \hat{\psi}^*(\hat{\mathbf{r}}_1) \hat{\psi}(\hat{\mathbf{r}}_2) \rangle$ of the radiation pattern. We conclude then that *the autocorrelation function of the field is uniquely determined everywhere outside the source volume by the autocorrelation function of the radiation pattern and vice versa.* It follows that the inverse problem for random sources reduces, in analogy to its deterministic counterpart, to that of deducing the autocorrelation function of the source from the autocorrelation function of the radiation pattern.

A more complete analogy between the deterministic and random inverse source problems is obtained by intro-

ducing the sixfold Fourier transform of the source autocorrelation function

$$\Phi(\mathbf{k}_1,\mathbf{k}_2) = \int d^3r_1 \int d^3r_2\, R_\rho(\mathbf{r}_1,\mathbf{r}_2) e^{-i(\mathbf{k}_2\cdot\mathbf{r}_2 - \mathbf{k}_1\cdot\mathbf{r}_1)} \quad (2.8)$$

This quantity is readily shown to be equal to the autocorrelation function of the Fourier transform of the source

$$\Phi(\mathbf{k}_1,\mathbf{k}_2) = \langle \tilde{\rho}^*(\mathbf{k}_1)\tilde{\rho}(\mathbf{k}_2)\rangle. \quad (2.9)$$

The inverse problem for random sources thus becomes that of deducing $\Phi(\mathbf{k}_1,\mathbf{k}_2)$ for all values of the wave vectors $\mathbf{k}_1,\mathbf{k}_2$ from its boundary value

$$\Phi(k_0\hat{\mathbf{r}}_1,k_0\hat{\mathbf{r}}_2) = \langle \hat{\psi}^*(\mathbf{r}_1)\hat{\psi}(\mathbf{r}_2)\rangle. \quad (2.10)$$

as given by the autocorrelation function of the radiation pattern.

3. NONUNIQUENESS IN THE INVERSE PROBLEM

In both the deterministic and random inverse source problems the data has lower *dimensionality* than what is required to uniquely specify the source in question. For example, in the deterministic case the data consists of the prescription of an unknown function $\tilde{\rho}(\mathbf{k})$ of *three* variables (say the Cartesian components of $\mathbf{k}$) on the *two-dimensional* surface $|\mathbf{k}| = k_0$. It is not possible to uniquely continue $\tilde{\rho}(\mathbf{k})$ from its boundary value over such a two-dimensional surface even for the class of sources considered here (localized and continuous) whose transforms are entire analytic functions of $\mathbf{k}$.[9] In particular, the entire function $\tilde{\rho}(\mathbf{k})$ is uniquely specified for all values of $\mathbf{k}$ if and only if it is specified over a finite *volume element* in $\mathbf{k}$ space,[10] i.e., over a three-dimensional region. The data in the deterministic inverse source problem specifies $\tilde{\rho}(\mathbf{k})$ only over a two-dimensional surface and, thus, is not sufficient to uniquely determine this quantity for all values of $\mathbf{k}$.

A similar situation prevails in the inverse problem for random sources. In this case the data is a prescription of the entire analytic function $\Phi(\mathbf{k}_1,\mathbf{k}_2)$ of *six* variables over the four dimensional surface $|\mathbf{k}_1| = |\mathbf{k}_2| = k_0$ while unique determination requires that it be specified over a finite volume element in $(\mathbf{k}_1,\mathbf{k}_2)$ space, i.e., over a *six-dimensional* region.

Examples of deterministic sources that are not uniquely specified from their values on the surface $|\mathbf{k}| = k_0$ are provided by the class of so-called *nonradiating sources*.[3] These sources are readily constructed by applying the operator $(\nabla^2 + k_0^2)$ to any thrice differentiable function $Q(\mathbf{r})$ localized within V. In particular the source

$$\rho_{\text{N.R.}}(\mathbf{r}) = (\nabla^2 + k_0^2)Q(\mathbf{r}) \quad (3.1)$$

possesses the Fourier transform

$$\tilde{\rho}_{\text{N.R.}}(\mathbf{k}) = \int d^3r[(\nabla^2 + k_0^2)Q(\mathbf{r})]e^{-i\mathbf{k}\cdot\mathbf{r}} = -(k^2 - k_0^2)\tilde{Q}(\mathbf{k}) \quad (3.2)$$

with $\tilde{Q}(\mathbf{k})$ being the Fourier transform of $Q(\mathbf{r})$. It is seen from Eq. (3.2) that irrespective of the choice of the function $Q(\mathbf{r})$ the source transform $\tilde{\rho}_{\text{N.R.}}(\mathbf{k})$ vanishes when $|\mathbf{k}| = k_0$. The class of nonradiating sources thus possess transforms which are identical on the surface $|\mathbf{k}| = k_0$ (namely, zero) but are rather arbitrary for values of $\mathbf{k}$ not on this surface.

The fact that a nonradiating source $\rho_{\text{N.R.}}(\mathbf{r})$ possesses a transform which vanishes when $|\mathbf{k}| = k_0$ means that it generates a field which vanishes identically outside its region of localization. It follows that the deterministic sources $\rho(\mathbf{r})$ and $\rho(\mathbf{r}) + \rho_{\text{N.R.}}(\mathbf{r})$ produce identical fields outside V and, hence, are both solutions to the same inverse source problem. The lack of uniqueness of solutions to the deterministic inverse source problem can thus be viewed as being due to the possible existence of nonradiating sources within the source volume.[11]

The lack of uniqueness of solutions to the random inverse source problem can likewise be be attributed to the possible existence of nonradiating sources within the source volume. In this case the function $Q(\mathbf{r})$ appearing in Eq. (3.1) is chosen to be a member of a random ensemble of thrice differentiable functions localized within V. The six-dimensional Fourier spectrum of the autocorrelation function of the random source

$$\rho'(\mathbf{r}) = \rho(\mathbf{r}) + \rho_{\text{N.R.}}(\mathbf{r}) \quad (3.3)$$

is found to be

$$\begin{aligned}\Phi'(\mathbf{k}_1,\mathbf{k}_2) &= \langle \tilde{\rho}^*(\mathbf{k}_1)\tilde{\rho}(\mathbf{k}_2)\rangle + \langle \tilde{\rho}^*(\mathbf{k}_1)\tilde{\rho}_{\text{N.R.}}(\mathbf{k}_2)\rangle \\ &\quad + \langle \tilde{\rho}^*_{\text{N.R.}}(\mathbf{k}_1)\tilde{\rho}(\mathbf{k}_2)\rangle + \langle \tilde{\rho}^*_{\text{N.R.}}(\mathbf{k}_1)\tilde{\rho}_{\text{N.R.}}(\mathbf{k}_2)\rangle \\ &= \Phi(\mathbf{k}_1,\mathbf{k}_2) - (k_2^2 - k_0^2)\langle \tilde{\rho}^*(\mathbf{k}_1)\tilde{Q}(\mathbf{k}_2)\rangle \\ &\quad - (k_1^2 - k_0^2)\langle \tilde{Q}^*(\mathbf{k}_1)\tilde{\rho}(\mathbf{k}_2)\rangle \\ &\quad + (k_1^2 - k_0^2)(k_2^2 - k_0^2)\langle \tilde{Q}^*(\mathbf{k}_1)\tilde{Q}(\mathbf{k}_2)\rangle. \end{aligned} \quad (3.4)$$

We conclude that

$$\Phi'(\mathbf{k}_1,\mathbf{k}_2)\big|_{|\mathbf{k}_1| = |\mathbf{k}_2| = k_0} = \Phi(\mathbf{k}_1,\mathbf{k}_2)\big|_{|\mathbf{k}_1| = |\mathbf{k}_2| = k_0} \quad (3.5)$$

independent of the statistical structure of the random process $Q(\mathbf{r})$. Thus, the two random sources $\rho(\mathbf{r})$ and $\rho'(\mathbf{r})$ generate fields possessing identical autocorrelation functions outside V and, thus, are both solutions to the same inverse problem.[12]

The discussion presented above assumes, of course, that the only information available concerning the unknown source is that generated by field measurements performed *exterior* to the source volume. If field measurements are allowed *internal* to the source region or if additional information is available which reduces the dimensionality of the source, then the inverse problem may possess a unique solution.

An example of the latter situation is provided by the class of *incoherent sources*. These sources possess autocorrelation functions of the form

$$R_\rho(\mathbf{r}_1,\mathbf{r}_2) = I(\mathbf{r}_1)\delta(\mathbf{r}_2 - \mathbf{r}_1), \quad (3.6)$$

where $I(\mathbf{r}_1)$ is a continuous, nonnegative function called the "intensity profile" of the source and $\delta(\mathbf{r}_2 - \mathbf{r}_1)$ is the three-dimensional Dirac delta function. The Fourier transform of $R_\rho(\mathbf{r}_1,\mathbf{r}_2)$ is found to be

$$\Phi(\mathbf{k}_1,\mathbf{k}_2) = \int_V d^3r_1\, I(\mathbf{r}_1)e^{-i(\mathbf{k}_2 - \mathbf{k}_1)\cdot\mathbf{r}_1} = \tilde{I}(\mathbf{k}_2 - \mathbf{k}_1). \quad (3.7)$$

It is seen from Eq. (3.7) that an incoherent source is characterized by an entire analytic function $\tilde{I}(\mathbf{k})$ of only *three variables*. Moreover,

$$\Phi(k_0\hat{\mathbf{r}}_1, k_0\hat{\mathbf{r}}_2) = \tilde{I}[k_0(\hat{\mathbf{r}}_2 - \hat{\mathbf{r}}_1)], \tag{3.8}$$

which shows that $\tilde{I}(\mathbf{k})$ is specified for all values of $\mathbf{k}$ lying within a sphere of radius $2k_0$ by the autocorrelation function of the radiation pattern. It follows that $\tilde{I}(\mathbf{k})$ is, in principle, completely determined (e.g., by analytic continuation) so that the inverse problem for incoherent sources admits a unique solution.

4. QUASIHOMOGENEOUS SOURCES

The concept of a quasihomogeneous source[7] is a natural generalization of the concept of statistical homogeneity to include sources of finite extent. The term "statistically homogeneous" is, of course, used to denote multidimensional, *stationary* random processes which, by definition, cannot be localized to a finite region. Carter and Wolf [7] argued, however, that many physical sources behave *locally* as though they were statistically homogeneous and thus have autocorrelation functions that can be approximated as follows:

$$R_\rho(\mathbf{r}_1, \mathbf{r}_2) = I[\tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)]g(\mathbf{r}_2 - \mathbf{r}_1). \tag{4.1}$$

The function $g(\mathbf{r})$ is a measure of the spatial coherence of the source and is assumed to be appreciably different from zero only for values of its argument $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ lying in some domain D which is much smaller than the source volume V. The quantity $I(\mathbf{R})$, called the source intensity, is assumed to be nonnegative, is a slowly varying function of its argument $\mathbf{R} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$ over the source volume V and vanishes outside this region. In addition, $I(\mathbf{R})$ is assumed to be essentially constant over the volume of coherence D. Sources having autocorrelation functions that can be approximated by the above model are called *quasihomogeneous.*

Substituting Eq. (4.1) into Eq. (2.8) yields the following expression for the sixfold Fourier transform of the autocorrelation function of a quasihomogeneous source:

$$\Phi(\mathbf{k}_1, \mathbf{k}_2) = \int d^3r_1 \int d^3r_2\, I[\tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)] \times g(\mathbf{r}_2 - \mathbf{r}_1) e^{-i(\mathbf{k}_2\cdot\mathbf{r}_2 - \mathbf{k}_1\cdot\mathbf{r}_1)}. \tag{4.2}$$

The above expression is simplified considerably by changing the variables of integration from $(\mathbf{r}_1, \mathbf{r}_2)$ to the $(\mathbf{R}, \mathbf{r})$ variables defined above. Performing this change of variables yields[13]

$$\Phi(\mathbf{k}_1, \mathbf{k}_2) = \int d^3R \int d^3r\, I(\mathbf{R}) g(\mathbf{r}) \times e^{-i\left\{(\mathbf{k}_2 - \mathbf{k}_1)\cdot\mathbf{R} + \left[\frac{\mathbf{k}_1 + \mathbf{k}_2}{2}\right]\cdot\mathbf{r}\right\}} = \tilde{I}(\mathbf{k})\tilde{g}(\mathbf{K}), \tag{4.3}$$

where we have introduced the wave vectors

$$\mathbf{k} = \mathbf{k}_2 - \mathbf{k}_1, \quad \mathbf{K} = \tfrac{1}{2}(\mathbf{k}_1 + \mathbf{k}_2) \tag{4.4}$$

and where $\tilde{I}(\mathbf{k})$ and $\tilde{g}(\mathbf{K})$ are the Fourier transforms of $I(\mathbf{R})$ and $g(\mathbf{r})$ respectively.

The data for the inverse problem specifies $\Phi(\mathbf{k}_1, \mathbf{k}_2)$ on the boundary $|\mathbf{k}_1| = |\mathbf{k}_2| = k_0$. It follows from Eqs. (4.4) that this surface corresponds to the boundary

$$k^2 + 4K^2 = 4k_0^2 \tag{4.5}$$

in $(\mathbf{k}, \mathbf{K})$ space. The inverse problem for quasihomogeneous sources can thus be stated as follows: *determine an intensity profile $I(\mathbf{R})$ and coherence function $g(\mathbf{r})$ having Fourier transforms $\tilde{I}(\mathbf{k})$ and $\tilde{g}(\mathbf{K})$, respectively, the product of which assumes a specified value (equal to the autocorrrelation function of the radiation pattern) for all values of $(\mathbf{k}, \mathbf{K})$ lying on the boundary defined in Eq.* (4.5).

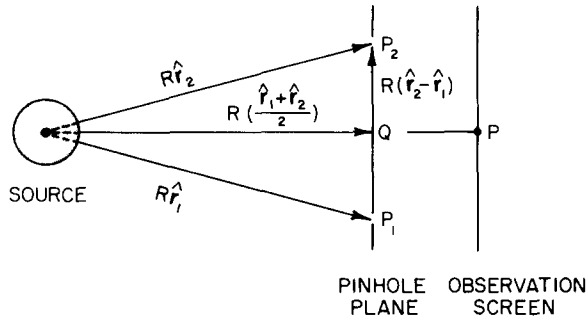

FIG. 1. Young's interference experiment. The autocorrelation function of the radiation pattern in the directions $\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2$ is determined by measuring the fringes in the vicinity of the point P on the observation screen.

The uniqueness question in the inverse problem for quasihomogeneous sources is complicated considerably by the requirements that the spectrum $\Phi(\mathbf{k}_1, \mathbf{k}_2)$ be factorizable in the form required by Eq. (4.3) and that the transforms $I(\mathbf{R})$ and $g(\mathbf{r})$ of the two factors $\tilde{I}(\mathbf{k})$ and $\tilde{g}(\mathbf{K})$, respectively, possess the properties described earlier. If, however, we restrict our attention to sources for which the two functions $\tilde{I}(\mathbf{k})$ and $\tilde{g}(\mathbf{K})$ are analytic and for which one of these two functions is known, then it is not difficult to show that the inverse problem admits a unique solution. For example, if $\tilde{I}(\mathbf{k})$ is known, then we conclude from the statement of the inverse problem given above that $\tilde{g}(\mathbf{K})$ can be determined over a *volume* element in $\mathbf{K}$ space, namely for all values of $\mathbf{K}$ lying within the sphere, centered at $\mathbf{K} = \mathbf{0}$, and of radius k_0. Analyticity of $\tilde{g}(\mathbf{K})$ then allows this quantity to be determined everywhere by analytic continuation.

The situation considered above where either the intensity profile $I(\mathbf{R})$ or coherence function $g(\mathbf{r})$ is known is of great interest both because it occurs frequently in practice and because for such cases the unknown source function [either $\tilde{I}(\mathbf{k})$ or $\tilde{g}(\mathbf{K})$] can be readily determined from far field interference experiments. To see this, let us consider a Young's interference experiment[14] performed on the surface of a sphere of radius R as illustrated in Fig. 1. The interference fringes observed in the vicinity of the central point P on the observation screen allows one to determine the spatial coherence function of the field at the points P_1, P_2 in the aperture plane. In the wave zone ($k_0 R \to \infty$) this coherence function is proportional to the autocorrelation function of the radiation pattern evaluated in the directions $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ so that from this experiment we determine[15]

$$\tilde{I}[k_0(\hat{\mathbf{r}}_2 - \hat{\mathbf{r}}_1)]\tilde{g}\left[k_0\left(\frac{\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2}{2}\right)\right] = \langle \hat{\psi}^*(\hat{\mathbf{r}}_1)\hat{\psi}(\hat{\mathbf{r}}_2)\rangle. \tag{4.6}$$

By varying the locations of the pinholes at points P_1, P_2 the wave vector $\mathbf{k} = k_0(\hat{\mathbf{r}}_2 - \hat{\mathbf{r}}_1)$ can be made to assume all values within a sphere of radius $2k_0$ while the wave vector

$\mathbf{K} = k_0(\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2)/2$ can be made to assume all values within a sphere of radius k_0. It follows that at the very least a band-limited approximation to one of the functions can be determined if the other is known. When applied to the problem of determining the intensity profile of a source of known coherence, the procedure is quite analogous to that used in determining the intensity profile of stellar sources by means of the Michelson stellar interferometer.[16] This latter procedure requires, however, that the stellar source be approximated by a *planar, incoherent* source so that the Van Cittert–Zernike theorem can be applied,[17] whereas in the procedure outlined above it is only necessary to assume that the stellar source can be approximated by a quasihomogeneous source of known coherence properties.

ACKNOWLEDGMENT

The author is indebted to Professor Emil Wolf for a number of valuable suggestions concerning the presentation of the research reported here.

[1]C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves* (Springer, New York, 1969), p. 339.

[2]A.J. Devaney and E. Wolf, J. Math. Phys. **15**, 234 (1974).

[3]An account of the early work in the theory of nonradiating sources is included in D. Bohm and M. Weinstein, Phys. Rev. **74**, 1789 (1948). More recent developments are reported by T. Erber and S.M. Prastein, Acta Phys. Austr. **32**, 224 (1970); J.B. Arnett and G.H. Goedecke, Phys. Rev. **168**, 1424 (1968); A.J. Devaney and E. Wolf, Phys. Rev. D **8**, 1044 (1973) and B.J. Hoenders, *Inverse Source Problems in Optics*, edited by H.P. Baltes (Springer-Verlag, Heidelberg, 1978) pp. 41–82. The importance of nonradiating sources to uniqueness in the deterministic inverse source problem is discussed by C. Müller, IRE Trans. Antennas Propag. **AP-4**, 3, 224 (1956) and by H. Bleistein and J.K. Cohen, J. Math. Phys. **18**, 194 (1977). The relevance of these sources to the inverse scattering problem is demonstrated in A.J. Devaney, J. Math. Phys. **19**, 1526 (1978).

[4]Throughout the paper ensemble averages will be denoted by angular brackets $\langle\ \rangle$ enclosing the quantity to be averaged and complex conjugates by a superscript asterisk * on the quantity being conjugated.

[5]In some applications the field $\psi(\mathbf{r})$ is the Fourier amplitude at frequency $\omega_0 = ck_0$ of a stationary time dependent field. For such cases the definition of the autocorrelation function $\Gamma(\mathbf{r}_1,\mathbf{r}_2)$ given in Eq. (1.5) has to be modified to include an integration over an infinitesmal band of frequencies surrounding ω_0 and the resulting quantity is referred to as the cross-spectral density function of the *time dependent* field. [See, for example, the discussion presented in E.W. Marchand and E. Wolf, J. Opt. Soc. Am. **62**, 379 (1972).]

[6]The quantity $\Phi(\mathbf{k}_1,\mathbf{k}_2)$ is, of course, simply the sixfold Fourier transform of the source autocorrelation function $R_p(\mathbf{r}_1,\mathbf{r}_2)$ and is, thus, the *generalized spectral density function* of the source random process. [See, for example, R.A. Silverman, IRE Trans. Inf. Theory **3**, 182 (1957).]

[7]Quasihomogeneous two-dimensional sources were introduced by W.H. Carter and E. Wolf [J. Opt. Soc. Am. **67**, 785 (1977)] and find many applications in the modern theory of radiometry [cf. E. Wolf and E. Collett, Opt. Comm. **25**, 293 (1978); E. Wolf, J. Opt. Soc, Am. **68**, 6 (1978)]. Three-dimensional quasihomogeneous sources and the properties of the fields they generate will be discussed in a forthcoming paper by E. Wolf and W.H. Carter.

[8]P.M. Morse and K.U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 7.

[9]The analyticity of $\tilde{\rho}(\mathbf{k})$ follows from a three-dimensional version of a well-known theorem that the Fourier transform of a continuous function which vanishes outside a finite interval is an entire analytic function. The multidimensional form of the theorem is the Plancherel-Polya theorem (cf. Ref. 10, p. 352).

[10]B.A. Fuks, *Introduction to the Theory of Analytic Functions of Several Complex Variables* (American Mathematics Society, Providence, Rhode Island, 1963).

[11]An analogous situation occurs in the inverse scattering problem where it is possible to have *nonscattering* potentials within the scattering volume (cf. the paper by this author quoted in Ref. 3).

[12]An account of the theory of random nonradiating sources will be presented in a forthcoming paper by B.J. Hoenders and H.P. Baltes.

[13]The procedure used in converting from the $(\mathbf{r}_1,\mathbf{r}_2)$ variables to the $(\mathbf{R},\mathbf{r})$ variables and the introduction of the frequency variables $(\mathbf{k},\mathbf{K})$ defined in Eq. (4.4) is entirely analogous to the treatment presented in Reference 7 for two-dimensional sources.

[14]M. Born and E. Wolf, *Principles of Optics*, 3rd ed. (Pergamon, New York, 1965), Sec. 7.2.

[15]The relationship (4.6) between the transforms $\tilde{I}$ and $\tilde{g}$ and the autocorrelation function of the radiation pattern is a three-dimensional version of the generalized Van Cittert–Zernike theorem (A. Walther, J. Opt. Soc. Am. **58**, 1256 (1968); E.W. Marchand and E. Wolf, J. Opt. Soc. Am. **62**, 379 (1972) as applied to quasihomogeneous planar sources (cf., the paper by Carter and Wolf quoted in Ref. 7). See also: H.P. Baltes and B. Steinle, Lett. Nuovo Cimento **18**, 313 (1977).

[16]See, for example, L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

[17]The requirement that the stellar source be incoherent can be dropped if the generalized Van Cittert–Zernike theorem (see Ref. 15) is used rather than the classical form of this theorem (cf. Ref. 14, Sec. 10.4.2).

Theory of strings and membranes in an external field. II. The string

A. Aurilia

Blackett Laboratory, Imperial College, London SW7 2BZ, England
and Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Italy[a)]

D. Christodoulou[b)]

Max Planck Institut für Physik und Astrophysik, München, Fed. Rep. Germany

(Received 6 February 1979)

We apply the general formulation of the theory of extended systems developed previously to the case of a relativistic string in an external field. This problem is relevant for the construction of models of strongly interacting particles. When the external field is static and uniform, explicit solutions are presented and classified according to their symmetry properties.

I. INTRODUCTION AND SUMMARY

In a previous paper,[1] hereafter referred to as I, we presented a general formulation of the theory of relativistic systems that are extended in space under the influence of an external field. Following that approach, in this article we study in detail the motion of a relativistic string in an external generalized Maxwell field. The relevance of the classical solutions to this problem for the physics of strongly interacting particles was emphasized in I. For instance, mesonic resonances are usually associated with the states of motion of strings with open ends and the amplitudes of the dual models can be reproduced by fusion and fission of such objects.[2] A closed string, carrying the same quantum numbers as the vacuum, has a special status in the string picture of hadrons and is most suitable for describing the nonresonant background of strong interactions[2,3] (the Pomeron). In the framework of relativistic field theory, the strings were also interpreted as the relativistic counterpart of the magnetic flux lines (vortices) found in the study of type II superconductors.[4,5] This latter interpretation of dual strings is based on a rather general and deep analogy between spontaneously broken gauge theories and the theory of superconductivity.[6] This observation suggests a physical picture of quark confinement in which an extended object is kept in equilibrium against the pressure of a surrounding charged superfluid.[4,6] In the field theoretical formulation these extended stringlike or baglike objects are viewed as domains embedded in a mass-inducing vacuum and constitute most suitable traps for the hadron constituents.

The above arguments are enough to motivate a deeper investigation of the classical foundation of the string theory, and this is the main purpose of this paper.

Our formulation is geometric: It solves in a natural way the problem of isolating the dynamical degrees of freedom of the string and applies equally well to extended objects of higher dimensionality[1,7] (membranes).

Let us recall from I that the evolution of the string is completely described by the following pair of dynamical equations

$$\rho\left(\frac{\partial u}{\partial t}+\nu w-N^2k\right)=cF_nN, \tag{1.1a}$$

$$\rho\left(\frac{\partial w}{\partial t}-\nu u\right)=cF_bN, \tag{1.1b}$$

where

$$F_n=-(1-u^2)E_b-uwE_n-Bw, \tag{1.2a}$$

$$F_b=(1-w^2)E_n+uwE_b+Bu. \tag{1.2b}$$

Here k is the curvature of the string, $\bar{l},\bar{n}$, and $\bar{b}$ are the unit tangent, normal, and binormal vectors respectively, and ν is the normal angular velocity of the osculating plane:

$$\nu=\bar{n}\cdot\frac{\partial\bar{b}}{\partial t}. \tag{1.3}$$

The velocity of the string is expressed as

$$\bar{v}=u\bar{n}+w\bar{b}, \tag{1.4}$$

and

$$N=(1-|\bar{v}|^2)^{1/2}. \tag{1.5}$$

Furthermore the dual 1-form of the external field is given by

$$*F=Bdx^0-\sum_{i=1}^{3}E^i dx^i. \tag{1.6}$$

Finally ρ and c are the generalized mass and generalized charge of the string, respectively. In this paper we discuss the case in which $*F$ is a given constant 1-form, corresponding to a uniform, static external field. This constant 1-form defines a family of parallel hyperplanes. If the normal to these hyperplanes is spacelike, $(*F)^2>0$ and we shall call the field "electric." It can be made purely electric by a suitable Lorentz transformation. If the normal is timelike, $(*F)^2<0$ and we shall call the field "magnetic." In the latter case $*F$ constitutes a Lorentz frame. In that frame the field is purely magnetic ($\bar{E}=0$). The energy of the string, defined only for a closed string, was found to be[1]

$$\mathscr{E}=\rho\int_{S^1}\frac{\mu_\gamma}{N}+c\int_\Sigma\bar{E}\cdot d\bar{\Sigma}=\rho\int_{S_1}\frac{\mu_\gamma}{N}+c\bar{m}\cdot\bar{E}. \tag{1.7}$$

Here μ_γ is the canonical measure of the metric γ induced on S^1, the topological model of the closed string and Σ is any surface having the string $\bar{x}(S^1)$ as boundary at a given time. Furthermore, we have expressed the potential energy of the

[a)]Permanent address.
[b)]Alexander von Humboldt fellow.

 0022-2488/79/081692-08$01.00

string in terms of the moment

$$\bar{m} = \tfrac{1}{2}\int_{x(S^1)} \bar{x}\times d\bar{l}. \tag{1.8}$$

In the following we shall study string configurations possessing a symmetry group. A "symmetry group" of the string is a subgroup P' of the Poincaré group, leaving the world track of the string invariant. In terms of the embedding $\xi: K\to R^4$, where $K\cong R\times M$ is the topological model of the world track (cf. I), we shall call ξ invariant under P' iff

$$\forall\, p'\in P' \quad \exists f\in\mathscr{D}^+(K); p\circ\xi = \xi\circ f,$$

where $\mathscr{D}^+(K)$ is the group of orientation preserving diffeomorphisms of K. The infinitesimal version of the above statement is the following: For every generator Λ of P' there is a vector field η in K such that

$$\eta^a \frac{\partial \xi^\mu}{\partial s^a} = \Lambda^\mu{}_\nu \xi^\nu. \tag{1.9}$$

We present explicit solutions which are invariant under translation either in time (static case, cf. Sec. II) or in space (straight lines, cf. Sec. IV). The knowledge of such classical solutions is an essential step towards the formulation of a quantized theory of interacting strings.

II. STATIC CASE

In the static case $u = w = v = 0$, $N = 1$ and the dynamical equations reduce to:

$$k = \frac{c}{\rho}\bar{E}\cdot\bar{b}, \quad \bar{E}\cdot\bar{n} = 0. \tag{2.1}$$

From the above equations it is evident that the general solution of the free problem ($c = 0$) is a straight line. Further, for $c\neq 0$ a straight line solution must be either parallel or antiparallel to E.

In the nonstraight line case, differentiating the first of Eqs. (2.1) with respect to the arc length s, taking into account (the third of) the Frenet formulas

$$\frac{d\bar{l}}{ds} = k\bar{n}, \quad \frac{d\bar{n}}{ds} = -k\bar{l} + t\bar{b}, \quad \frac{d\bar{b}}{ds} = -t\bar{n}, \tag{2.2}$$

where t is the torsion of the curve

$$t = \frac{1}{k^2}\left\|\frac{d\bar{x}}{ds}, \frac{d^2\bar{x}}{ds^2}, \frac{d^3\bar{x}}{ds^3}\right\|, \tag{2.3}$$

and using the second of Eqs. (2.1), we obtain

$$\frac{dk}{ds} = 0.$$

Likewise differentiating the second of Eqs. (2.1) with respect to s and using the second of (2.2) and the first of (2.1) we obtain

$$t = (c/\rho)\bar{E}\cdot\bar{l}. \tag{2.4}$$

Finally, differentiating the above equation with respect to s and using the first of (2.2) and the second of (2.1) we get

$$\frac{dt}{ds} = 0. \tag{2.5}$$

Thus, the curve has constant curvature and torsion. Hence, it is a circular helix with axis in the direction of $\bar{E}$ [the latter follows from the second of (2.1)] with:

$$k^2 + t^2 = (c/\rho)^2(E_b^2 + E_l^2) = (cE/\rho)^2, \tag{2.6}$$

where $E = |\bar{E}|$. The radius of the helix (a) and the period divided by $2\pi(\gamma)$, are expressed in terms of k and t by:

$$a = k/(k^2 + t^2), \quad \gamma = t/(k^2 + t^2).$$

From Eq. (2.6) we obtain the following relation between a and γ:

$$\gamma^2 = \frac{1}{(cE/\rho)^2} - a^2. \tag{2.7}$$

Thus $0\leqslant a\leqslant 1/|cE/\rho|$. The two limiting cases are the straight line in the direction of $\bar{E}$ ($a = k = 0$) and the circle of radius $a = 1/(cE/\rho)$ ($\gamma = t = 0$).

The helix is right-handed if $t > 0$ and left-handed if $t < 0$.

III. SYMMETRY OF A SPACE TRANSLATION [T(1) DYNAMIC]

With this symmetry, the topological model of the string is R^1 and $\bar{x}(R^1)$ is a straight line ($k = 0$) and the triad $(\bar{l},\bar{n},\bar{b})$ is constant ($v = 0$).

A. Electric case

We can always choose $\bar{n}$ such that $E_b = 0$. With this choice, the dynamical equations reduce to:

$$\dot{u} = -\beta_1 uwN, \quad \dot{w} = \beta_1(1 - w^2)N, \tag{3.1}$$

where $\beta_1 = cE_n/\rho$. If $\beta_1 = 0$, Eqs. (3.1) imply $u = \gamma_1$, $w = \gamma_2$ (constants). Thus we have motion with constant velocity, which can be brought to rest by a suitable Lorentz transformation. If $\beta_1 \neq 0$, the first integral of the above equations is

$$\frac{u}{N} = \gamma, \quad \frac{w}{N} = \beta_1 t, \quad \text{(up to a time translation)},$$

which, setting $\dot{a} = u$, $\dot{b} = w$ (a, b distance traveled in $\bar{n}$ and $\bar{b}$ direction respectively), may in turn be integrated to give:

$$a = \frac{\gamma}{\beta_1}\log|\beta_1 t + (1 + \gamma^2 + \beta_1^2 t^2)^{1/2}|,$$

$$b^2 - t^2 = \frac{1+\gamma^2}{\beta_1^2}.$$

These equations describe motion with constant proper acceleration in the direction $\bar{b}$:

$$\alpha^\mu = \frac{1}{N}\frac{\partial}{\partial t}\left(\frac{1}{N}\frac{\partial x^\mu}{\partial t}\right), \quad \alpha^2 = \beta_1^2(1 + \gamma^2).$$

B. Magnetic case

The dynamical equations in this case read

$$\dot{u} = \lambda N w, \quad \dot{w} = -\lambda N u, \tag{3.2}$$

where $\lambda = cB/\rho$. A first integral of the above equations is $N = (1 - \beta^2)^{1/2}$ [$\beta = (u^2 + w^2)^{1/2}$ a constant, $0\leqslant\beta\leqslant 1$]. Then Eqs. (3.2) can be integrated to give:

$$\dot{a} = u = \beta\sin[\lambda(1 - \beta^2)^{1/2}t],$$

$$\dot{b} = w = \beta\cos[\lambda(1 - \beta^2)^{1/2}t]$$

(up to an arbitray time translation). The above equations may in turn be integrated to obtain:

$$a = -\frac{\beta}{\lambda(1-\beta^2)^{1/2}}\cos[\lambda(1-\beta^2)^{1/2}t],$$

$$b = \frac{\beta}{\lambda(1-\beta^2)^{1/2}}\sin[\lambda(1-\beta^2)^{1/2}t]$$

[up to an arbitrary translation in the $(\bar{n},\bar{b})$ plane]. We conclude that the string describes circular orbits on its normal plane, of radius

$$r = \frac{1}{|\lambda|}\frac{\beta}{(1-\beta^2)^{1/2}} \quad \left(\beta = \frac{r|\lambda|}{\sqrt{1+r^2\lambda^2}}\right)$$

and with frequency (counterclockwise for $\lambda > 0$, clockwise for $\lambda < 0$)

$$\omega = |\lambda|(1-\beta^2)^{1/2} = |\lambda|/(1+r^2\lambda^2)^{1/2}.$$

IV. AXIAL SYMMETRY [SO(2) DYNAMIC]

With this symmetry, the topological model of the string is S^1 and $\bar{x}(S^1)$ is a circle of radius a, and $\bar{b}$ is constant. Thus, $k = 1/a$ and $\nu = 0$. We define b to be the distance traveled in the $\bar{b}$ direction.

A. Electric case

Consistency with the symmetry requires that the electric field $\bar{E}$ be normal to the plane of the circles

$$\bar{E} = E\bar{b}.$$

The dynamical equations (1.1) reduce to

$$\dot{u} = \frac{N^2}{a} - \lambda N(1-u^2),$$

$$\dot{w} = \lambda N u w, \quad N = (1-u^2-w^2)^{1/2} \tag{4.1}$$

with $\dot{a} = -u, \dot{b} = w$, and $\lambda = cE/\rho$. A first integral of the above equations is the energy integral (1.7). In the case we are considering,

$$\int_{S^1}\mu_\gamma/N = 2\pi a/N, \quad \bar{m} = \pi a^2\bar{b},$$

and the energy integral has the form

$$2a/N - \lambda a^2 = \mathscr{E}/\pi\rho. \tag{4.2}$$

If $\lambda = 0$, Eq. (4.1) implies that $w = \text{const}$, and we can set $w = 0$ by a suitable Lorentz transformation. Then, Eq. (4.2) reads

$$N = (1-\dot{a}^2)^{1/2} = a/a_0, \tag{4.3}$$

where we have defined $a_0 = \mathscr{E}/2\pi\rho$. The solution of this equation is

$$a = a_0\cos(t/a_0). \tag{4.4}$$

This solution describes a circular string which starts from rest at $t = 0$, with radius $a = a_0$, and contracts monotonically to the singularity $a = 0$, at time $T = (\pi/2)a_0$. The solution is tangent to the null cone at the singularity ($N \to 0$ as $a \to 0$).

If now $\lambda \neq 0$, defining a parameter $\beta = \mathscr{E}/\pi\rho\lambda$, we may write Eq. (4.2) in the form

$$N = \frac{2}{\lambda}\left(\frac{a}{a^2+\beta}\right). \tag{4.5}$$

In the second of Eqs. (4.1) we may express dw/dt as $-udw/da$, thus obtaining

$$\frac{dw}{da} = -2\lambda N w. \tag{4.6}$$

Since N is known as a function of a from Eq. (4.5), we can integrate the above equation to obtain

$$w = \gamma\exp\left(-2\lambda\int N da\right) = \gamma/|a^2+\beta|. \tag{4.7}$$

Consequently, we have

$$u^2 = 1 - N^2 - w^2 = y(x)/(x+\beta)^2, \tag{4.8}$$

where $x = a^2$, and $y(x)$ is the quadratic polynomial

$$y(x) = x^2 - 2(2/\lambda^2 - \beta)x + \beta^2 - \gamma^2. \tag{4.9}$$

Finally, if N given by Eq. (4.5) and u given by (4.8) are inserted in the first of Eqs. (4.1), that equation is satisfied.

We begin the discussion of the solutions by first noting the following general properties: All solutions which reach the singularity $a = 0$ are tangent to the light cone at the singular point; $N \to 0$ as $a \to 0$. In addition, all unbounded solutions are asymptotically tangent to the light cone; $N \to 0$ as $a \to \infty$. As $a \to 0$, $u^2 \to 1 - \gamma^2/\beta^2$ and $du/da \to 0$. As $a \to \infty$, $u \to 0$.

If u does not change sign in the interval $[a_1, a_2]$, then

$$t(a_2) - t(a_1) = \int_{a_1}^{a_2} -\frac{da}{u} = \pm\int_{a_1}^{a_2}\frac{|a^2+\beta|}{[y(a^2)]^{1/2}}da,$$

and

$$b(a_2) - b(a_1) = \int_{a_1}^{a_2} -\frac{w da}{u} = \pm\int_{a_1}^{a_2}\frac{\gamma}{[y(a_2)]^{1/2}}da,$$

where we have $+$ or $-$ according to whether $u \leqslant 0$ or $\geqslant 0$ in the interval $[a_1, a_2]$, respectively.

The behavior of the solutions depends on the discriminant Δ of the quadratic polynomial $y(x)$:

$$\Delta = \left(\frac{2}{\lambda^2} - \beta\right)^2 + \gamma^2 - \beta^2 = \frac{4}{\lambda^2}\left(\frac{1}{\lambda^2} - \beta\right) + \gamma^2. \tag{4.10}$$

If $\Delta < 0$, $y(x)$ is positive definite and we have unbounded singular solutions (a ranges from 0 to ∞) with two branches corresponding to $u > 0$ and $u < 0$. Further, $t(a) - t(0)$, the time lapse between radius a and the singularity, is finite, and $b(\infty) - b(0)$, the total distance traveled in the x^3 direction, is also finite.

If $\Delta = 0$, the polynomial $y(x)$ has the double root

$$x_c = 1/\lambda^2 - \lambda^2\gamma^2/4. \tag{4.11}$$

For $x_c > 0$, we have the separatrices S_+ and S_- and their intersection, which is the critical point: $a = a_c = [x_c]^{1/2}$, $u = 0$. The separatrices S_+ and S_- are solutions for which $a \to a_c$ as $t \to +\infty$ and $-\infty$ respectively.

For $x_c = 0$, we have an unbounded asymptotically singular solution with two branches corresponding to $u \geqslant 0$ and $u \leqslant 0$, for which $t(a) - t(a_0) \to +\infty$ and $-\infty$ respectively, as $a \to 0$ [a_0 fixed $\neq 0$]. Also, $|b(a) - b(a_0)| \to \infty$ as $a \to 0$.

For $x_c < 0$, we have unbounded singular solutions, with the same properties as those stated for the case $\Delta < 0$.

If finally $\Delta > 0$, the polynomial $y(x)$ has two distinct roots

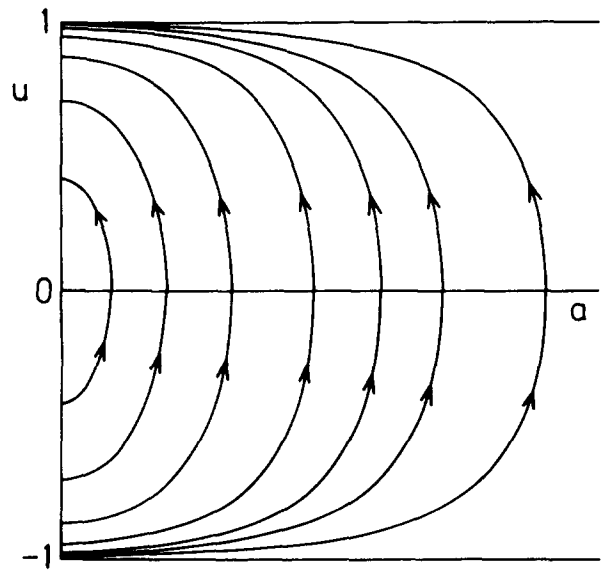

FIG. 1. The (a,u) diagram for $\lambda<0$.

$$x_{\pm}=(2/\lambda^2-\beta)\pm\sqrt{\Delta},\quad x_-<x_+. \tag{4.12}$$

For $x_-, x_+>0$ we have solutions with $a\leqslant(x_-)^{1/2}$ and $a\geqslant(x_+)^{1/2}$, which are bounded singular and unbounded regular, respectively. For the unbounded regular solutions, $b(\infty)-b[(x_+)^{1/2}]$ is finite. For $x_-\leqslant 0, x_+>0$ we have only the unbounded regular solutions $a\geqslant(x_+)^{1/2}$.

For $x_+, =0$ we have unbounded asymptotically singular solutions for which $|t(a)-t(a_0)|, |b(a)-b(a_0)|\to\infty$ as $a\to 0$.

Finally, for $x_+<0$, we have unbounded singular solutions.

In the case $\gamma=0$, we have $w=0$ and the motion of the string is confined to its plane. In the following we shall discuss the case $\gamma\neq 0$.

Case $\lambda<0$: In this case, Eq. (4.5) implies: $\beta<0, a^2<|\beta|$ (by definition $N>0$). The discriminant (4.10) is positive and among the roots of $y(x)$, $x_+>|\beta|$ and $x_-<|\beta|$. Further, $x_->0$ iff $|\beta|>|\gamma|$. Thus, we have the bounded singular solutions $a\leqslant(x_-)^{1/2}$ for $|\beta|>|\gamma|$ (Fig. 1).

Case $\lambda>0$: This case divides itself into three subcases: $x_c>0, x_c=0$, and $x_c<0$ [x_c defined by (4.11)] for which $0<|\gamma|<2/\lambda^2$, $|\gamma|=2/\lambda^2$, and $|\gamma|>2/\lambda^2$, respectively.

Subcase $0<|\gamma|<2/\lambda^2$ (Fig. 2): In this subcase, for $\beta=1/\lambda^2+\lambda^2\gamma^2/4$, which corresponds to $\Delta=0$, we have the separatrices

$$S_{\pm}: u=\pm\frac{(a^2-a_c^2)}{(a^2+1/\lambda^2+\lambda^2\gamma^2/4)} \tag{4.13}$$

and their intersection, which is the saddle-type critical point $a=a_c=(x_c)^{1/2}$, $u=0$. The solutions which correspond to the separatrices are (up to a time translation).

$$\mp t=a+\frac{1}{\lambda^2}\frac{1}{a_c}\log\left|\frac{a-a_c}{a+a_c}\right|,$$
$$\mp b=\frac{\gamma}{2a_c}\log\left|\frac{a-a_c}{a+a_c}\right|. \tag{4.14}$$

The solution which corresponds to the critical point represents a circle of constant radius a_c traveling in the direction normal to its plane with velocity $w=\gamma\lambda^2/2$. For $\gamma=0$ we obtain the static circle with radius $a_c=1/\lambda$ which we encountered in the study of the static problem.

For $\beta>1/\lambda^2+\lambda^2\gamma^2/4$ which corresponds to $\Delta<0$ we have the aforementioned unbounded singular solutions. These solutions have a minimum of u^2 at

$$a^2=\beta-\lambda^2\gamma^2/2, \tag{4.15}$$

which is given by

$$u^2_{\min}=1-\frac{1}{\lambda^2(\beta-\lambda^2\gamma^2/4)}. \tag{4.16}$$

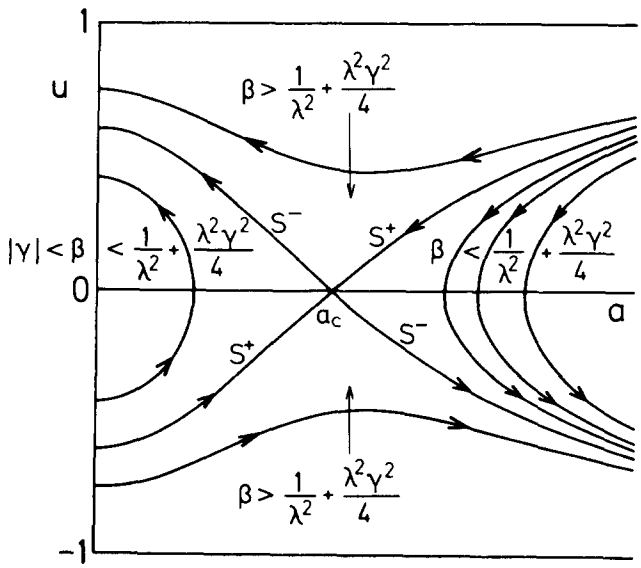

FIG. 2. The (a,u) diagram for $\lambda>0$, $0<|\gamma|<2/\lambda^2$.

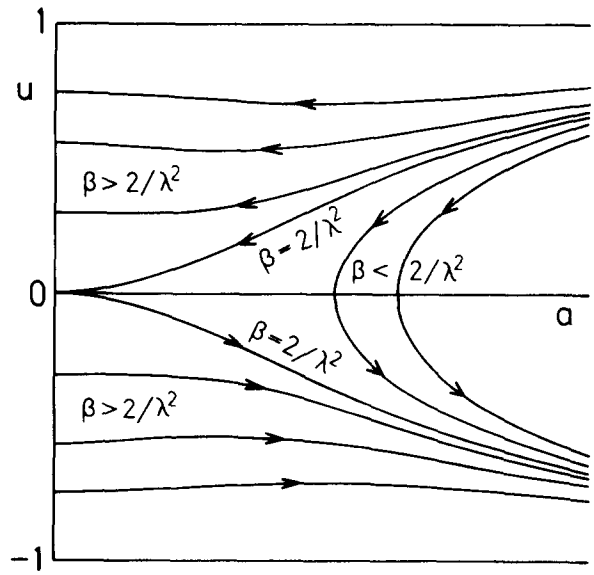

FIG. 3. The (a,u) diagram for $\lambda>0$, $|\gamma|=2/\lambda^2$.

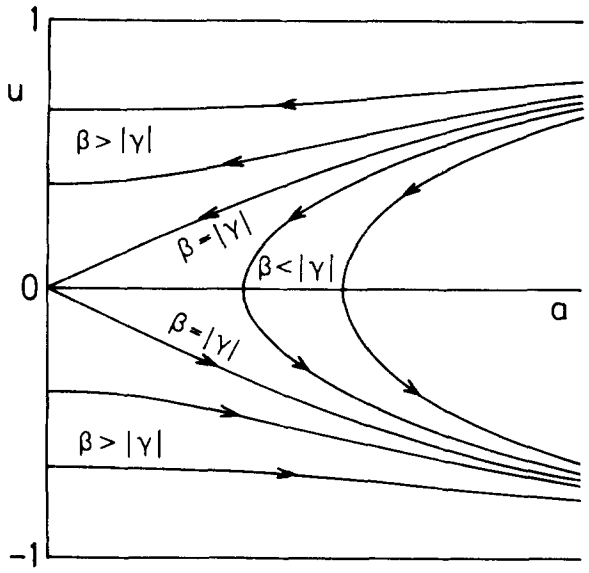

FIG. 4. The (a,u) diagram for $\lambda > 0$, $|\gamma| > 2/\lambda^2$.

Also, if $\beta < 2/\lambda^2$, $|db/da|$ has a maximum at $a^2 = 2/\lambda^2 - \beta < a_c^2$.

For $|\gamma| < \beta < 1/\lambda^2 + \lambda^2\gamma^2/4$ $(\Delta > 0)$, we have $0 < x_- < x_c$ and $x_c < x_+ < 2(2/\lambda^2 - |\gamma|)$. Thus, we have the pair of solutions $a \leqslant (x_-)^{1/2}$ and $a \geqslant (x_+)^{1/2}$, bounded singular and unbounded regular, respectively.

For $\beta \leqslant |\gamma|$ we either have $x_- \leqslant 0$ for $-|\gamma| \leqslant \beta \leqslant |\gamma|$ or $x_- < |\beta|$ for $\beta < -|\gamma|$. Thus, we have only the unbounded regular solution $a \geqslant (x_+)^{1/2} \geqslant 2(2/\lambda^2 - |\gamma|)$ {since for $\beta < 0, a^2 < |\beta|$ implies $N < 0$ [Eq. (4.5)]}.

Subcase $|\gamma| = 2/\lambda^2$ (Fig. 3): In this subcase, for $\beta = 2/\lambda^2$, which corresponds to $\Delta = 0$ with $x_c = 0$, we have the unbounded asymptotically singular pair of solutions

$$u = \pm a^2/(a^2 + 2/\lambda^2), \tag{4.17}$$

for which (up to a time translation)

$$\mp t = a - 2/\lambda^2 a, \quad \mp b = -\gamma/a. \tag{4.18}$$

For $\beta > 2/\lambda^2$ $(\Delta < 0)$, we have bounded singular solutions with a minimum of u^2 given by Eqs. (4.15) and (4.16), and with no maximum of $|db/da|$.

For $\beta < 2/\lambda^2$ $(\Delta > 0)$, we have the unbounded regular solutions $a \geqslant (x_+)^{1/2}$.

Subcase $|\gamma| > 2/\lambda^2$ (Fig. 4): In this subcase, for $\beta > |\gamma|$ we have unbounded singular solutions (For $\beta > 1/\lambda^2 + \lambda^2\gamma^2/4$, $\Delta > 0$; for $\beta = 1/\lambda^2 + \lambda^2\gamma^2/4$, $\Delta = 0$ and $x_c < 0$; for $1/\lambda^2 + \lambda^2\gamma^2/4 > \beta > |\gamma|$, $\Delta > 0$ and $x_+ < 0$.) These solutions possess a minimum of u^2 [given by Eqs. (4.15) and (4.16)] if $\beta > \lambda^2\gamma^2/2$ (no maximum of $|db/da|$).

For $\beta = |\gamma|$, $x_+ = 0$ and we have the unbounded asymptotically singular solutions

$$u^2 = \frac{a^2[a^2 + 2(|\gamma| - 2/\lambda^2)]}{(a^2 + |\gamma|)^2}, \tag{4.19}$$

for which (up to a time and x^3 translation

$$\mp t = [a^2 + 2(|\gamma| - 2/\lambda^2)]^{1/2} - \frac{|\gamma|}{[2(|\gamma| - 2/\lambda^2)]^{1/2}} \times \log\left| \frac{[2(|\gamma| - 2/\lambda^2)]^{1/2} + [a^2 + 2(|\gamma| - 2/\lambda^2)]^{1/2}}{a} \right|, \tag{4.20}$$

$$\mp b = -\frac{\gamma}{[2(|\gamma| - 2/\lambda^2)]^{1/2}} \times \log\left| \frac{[2(|\gamma| - 2/\lambda^2)]^{1/2} + [a^2 + 2(|\gamma| - 2/\lambda^2)]^{1/2}}{a} \right|.$$

Finally, for $\beta < |\gamma|$ we have the unbounded regular solutions $a \geqslant (x_+)^{1/2}$ $(x_+ > 0)$.

For all bounded singular solutions, as a decreases from $(x_-)^{1/2}$ to 0, t and b as functions of a can be expressed in the following elliptic integral form:

$$t = I(1) - I(a/\sqrt{x_-}), \quad b = J(1) - J(a/\sqrt{x_-}),$$

where

$$I(\xi) = (\lambda/|\lambda|\sqrt{x_+})[(\beta + x_+)F(k,\xi) - x_+ E(k,\xi)],$$

$$J(\xi) = (1/\sqrt{x_+})F(k,\xi),$$

with

$$k = \sqrt{x_-/x_+}.$$

The five qualitative types of solutions described above are displayed pictorially in Fig. 5.

B. Magnetic case

In this case the dynamical equations become

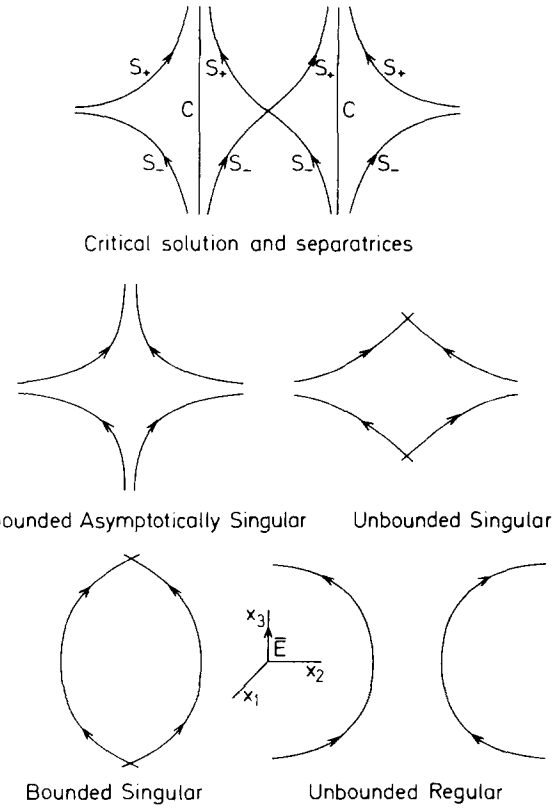

FIG. 5. The five types of solutions of the SO(2) problem (Electric case, $\gamma > 0$).

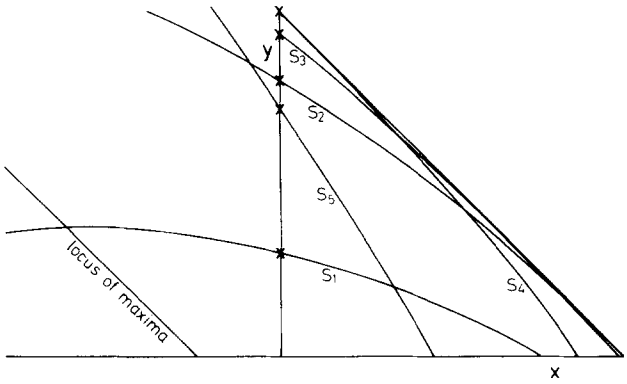

FIG. 6. $|\beta| < 1$.

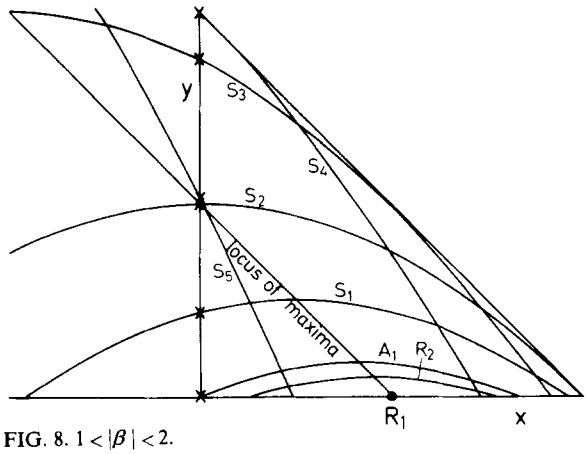

FIG. 8. $1 < |\beta| < 2$.

$$\dot{a} = -u, \quad \dot{b} = w, \tag{4.21}$$
$$\dot{u} = N^2/a - \lambda N w, \quad \dot{w} = \lambda N u,$$

where $N = [1 - (u^2 + w^2)]^{1/2}$.

A first integral of the above equations is the energy integral (1.7) which yields

$$N = a/a_0 \tag{4.22}$$

where

$$a_0 = \mathscr{E}/2\pi\rho > 0.$$

Equation (4.22) allows us to express w in terms of u and a:

$$w = \eta[1 - (a/a_0^2) - u^2]^{1/2}: \quad \eta = \pm 1. \tag{4.23}$$

Substituting for N and w from (4.22) and (4.23) respectively, in the third of Eqs. (4.21) we obtain the following pair of equations for the flow of a vector field X on the (a,u) plane (which is actually the $R^+ \times]-1,1[$ open subset of the plane)

$$\dot{a} = -u = X_a, \tag{4.24}$$
$$\dot{u} = a/a_0\{1/a_0 - \eta\lambda\,[1 - (a/a_0)^2 - u^2]^{1/2}\}. = X_u.$$

The critical point of this vector field ($X = 0$) is the point

$$u = 0, \quad a/a_0 = [1 - 1/\lambda^2 a_0^2]^{1/2}, \tag{4.25}$$

which exists iff $\eta\lambda > 0$ and $|\lambda a_0| > 1$. The matrix X' at the critical point is given by

$$(X')_{\text{crit}} = \begin{bmatrix} \dfrac{\partial X_a}{\partial a} & \dfrac{\partial X_a}{\partial u} \\ \dfrac{\partial X_u}{\partial a} & \dfrac{\partial X_u}{\partial u} \end{bmatrix}_{\text{crit}} = \begin{bmatrix} 0 & -1 \\ \lambda^2 - \dfrac{1}{a_0^2} & 0 \end{bmatrix}. \tag{4.26}$$

This matrix has imaginary eigenvalues

$$\mu_{\pm} = \pm i(\lambda^2 - 1/a_0^2)^{1/2} \quad \text{(characteristic frequencies)}. \tag{4.27}$$

Therefore, the critical point is a vortex point.

Multiplying the right-hand side of the second of Eqs. (4.24) by $-u$ and the left-hand side by $\dot{a}$ we obtain

$$-udu = a/a_0\{1/a_0 - \eta\lambda\,[1 - (a/a_0)^2 - u^2]^{1/2}\}da. \tag{4.28}$$

Let us now define new variables x and z by

$$x = (a/a_0)^2, \quad z = 1 - (a/a_0)^2 - u^2. \tag{4.29}$$

In tems of the new variables, Eq. (4.28) assumes the form

$$dz + \eta\beta(z)^{1/2}dx = 0, \tag{4.30}$$

where $\beta = \lambda a_0 \neq 0$. The above equation can be integrated to give

$$\eta\sqrt{z} = w = (1/2\beta)(\gamma - \beta^2 x), \tag{4.31}$$
$$z = 1 - x - y = (1/4\beta^2)(\gamma - \beta^2 x)^2,$$

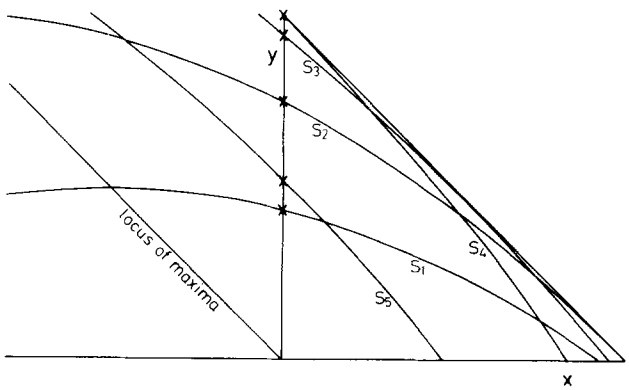

FIG. 7. $|\beta| = 1$.

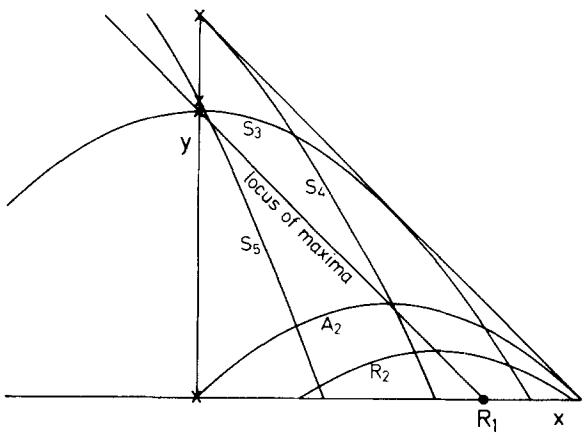

FIG. 9. $|\beta| = 2$.

where γ is a constant of integration and $y = u^2$. The above pair of equations constitute the complete first integral of Eqs. (4.21).

The physical region in the (x,y) plane is the right triangle defined by the conditions $x > 0, y \geqslant 0$ and $w^2 = 1 - x - y \geqslant 0$. The curves $y(x)$ are parabolas convex to the x axis, whose curvature depends only on β,

$$y = (\beta^2/4)(x_+ - x)(x - x_-). \tag{4.32}$$

Here, the roots $x_\pm$ are given by

$$x_\pm = \frac{2}{\beta^2}\left[\left(\frac{\gamma}{2} - 1\right) \pm \sqrt{\Delta}\right], \tag{4.33}$$

$(x_+ \geqslant x_-)$ and the discriminant Δ is

$$\Delta = 1 - \gamma + \beta^2. \tag{4.34}$$

For a fixed value of β the locus of maxima of the parabolas (4.32) is a straight line parallel to the hypotenuse of the right triangle which represents the physical region

$$x_1 = \frac{1}{\beta^2}(\gamma - 2), \quad y_1 = \frac{1}{\beta^2}(1 - \gamma + \beta^2), \tag{4.35}$$

or $x_1 + y_1 = 1 - \dfrac{1}{\beta^2}$.

The locus of maxima intersects the physical region iff $|\beta| > 1$. For $\gamma = \beta^2 + 1$, the discriminant Δ vanishes and the intersection of the corresponding parabola with the physical region (which is nonempty only for $|\beta| > 1$) is the critical point

$$x_+ = x_- = x_1 = 1 - 1/\beta^2, \quad y_1 = 0.$$

This is a solution which represents a circle of constant radius $a/a_0 = [1 - 1/\beta^2]^{1/2}$ traveling in a direction normal to its plane with velocity $w = 1/\beta$.

A parabola intersects the physical region, and therefore is relevant, if and only if $x_+ > 0$ and $y_1 \geqslant 0$. These conditions yield the folowing range of γ for a given β:

$$\begin{aligned} &|\beta| > 1: \quad -2|\beta| < \gamma \leqslant \beta^2 + 1, \\ &|\beta| \leqslant 1: \quad -2|\beta| < \gamma < 2|\beta|. \end{aligned} \tag{4.36}$$

A parabola is tangent to the straight line $x + y = 1$, of which the hypotenuse of the physical triangle is an interval, at

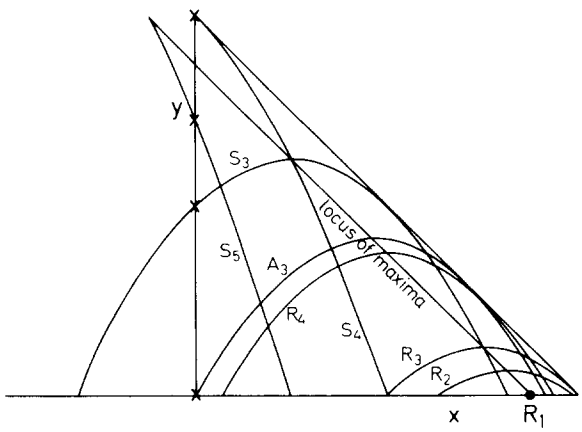

FIG. 10. $|\beta| > 2$.

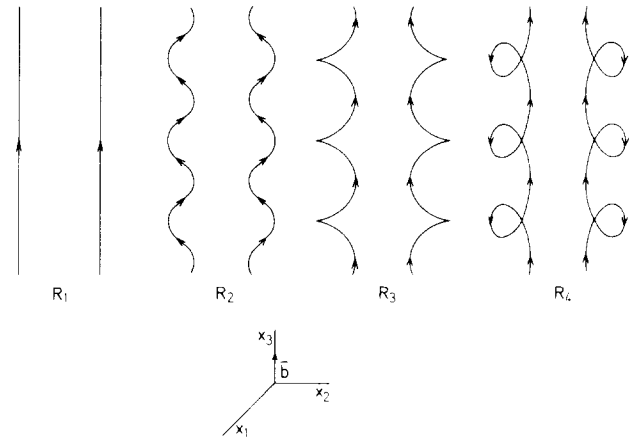

FIG. 11. The regular solutions. [Here $\beta > 0$, for $\beta < 0$ reverse direction of $\bar{b}$.]

$$x_0 = \gamma/\beta^2. \tag{4.37}$$

In terms of x_0, the first of Eqs. (4.31) may be expressed as

$$w = \tfrac{1}{2}\beta(x_0 - x), \tag{4.38}$$

from which it follows that βw is > 0, $= 0$, or < 0 whenever x is $< x_0$, $= x$, and $> x_0$, respectively.

In conclusion, we have a family of solutions of Eqs. (4.21), parametrized by two variables β and γ up to time translations and translations in the direction of $\bar{b}$. For any given pair (β,γ) all the properties of the solution can be read from the corresponding parabola (4.32) in the (x,y) plane.

Thus, we can distinguish three classes of solutions: (1) R: regular, (2) A: asymptotically singular, and (3) S: singular. The regular solutions correspond to $x_- > 0$ which is equivalent to $\gamma > 2|\beta|$. The asymptotically singular solutions are those for which $x_- = 0$, which is equivalent to $\gamma = 2|\beta|$. These solutions have the property that $x \to 0$ as $t \to \infty$, from which their name is motivated. Finally, the singular solutions are those for which $x_- < 0$, or, equivalently, $\gamma < 2|\beta|$. The regular and asymptotically singular solutions exist only for $|\beta| > 1$, as is evident from (4.36).

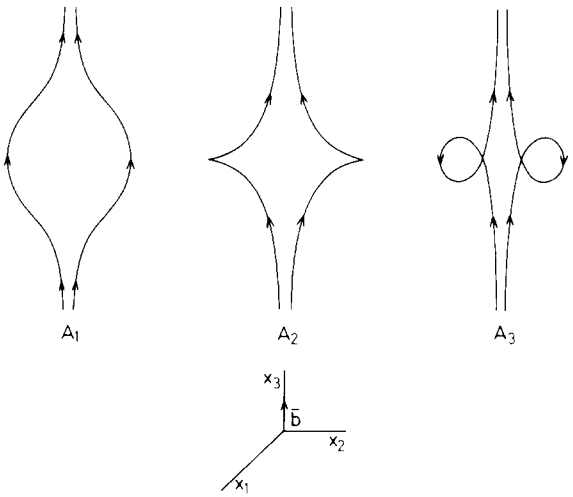

FIG. 12. The asymptotically singular solutions. [Here $\beta > 0$, for $\beta < 0$ reverse $\bar{b}$.]

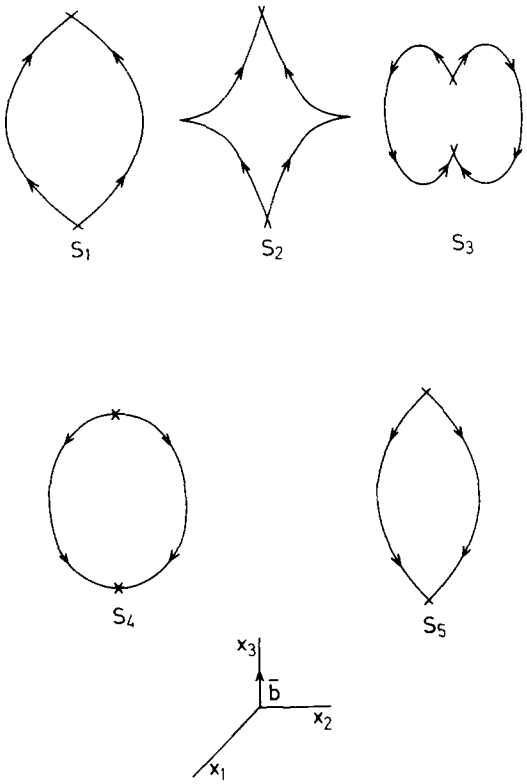

FIG. 13. The singular solutions. [Here $\beta > 0$, for $\beta < 0$ reverse $\bar{b}$.]

The regular solutions are of the following four types:
$(R_1)\gamma = \beta^2 + 1$: critical point,
$(R_2)\beta^2 < \gamma < \beta^2 + 1 : x_0 > 1$,
$(R_3)\gamma = \beta^2 : x_0 = 1$,
$(R_4)\gamma = \beta^2 : x_0 < 1$ [types (R_3) and (R_4) exist only for $|\beta| > 2$].
The asymptotically singular solutions are of the following three types:
$(A_1)|\beta| < 2 : x_0 > 1$,
$(A_2)|\beta| = 2 : x_0 = 1$,
$(A_3)|\beta| < 2 : x_0 < 1$.
Finally, the singular solutions are of the following five types:
$(S_1)\gamma > \beta^2 : x_0 > 1$,
$(S_2)\gamma = \beta^2 : x_0 = 1$,
$(S_3)0 < \gamma < \beta^2 : 0 < x_0 < 1$,
$(S_4)\gamma = 0 : x_0 = 0$,
$(S_5)\gamma < 0 : x_0 < 0$ [types (S_1) and (S_2) exist only for $|\beta| < 2$].
The solutions (R_3),(A_2), and (S_2) are at rest $(u = w = 0)$ at $x_+ = x_0 = 1$, with $db/da = 0$. The solution (S_4) starts and terminates at $x = 0$ with $u = -1$ and $+1$ respectively, and $w = 0$. (Figs. 6–13).

In the case of the regular solutions $x_- \leqslant x \leqslant x_+$ and as x decreases from x_+ to x_-, t and b as functions of a are expressed in the following elliptic integral form:

$$t = \frac{2a_0}{|\beta|\sqrt{x_+}} F(k,\xi),$$
$$b = \frac{\beta/|\beta|}{\sqrt{x_+}} [x_0 F(k,\xi) - x_+ E(k,\xi)], \tag{4.39}$$

up to a time and x_3 translation, where

$$k = \left(\frac{x_+ - x_-}{x_+}\right)^{1/2}, \quad \xi = \left(\frac{x_+ - x}{x_+ - x_-}\right)^{1/2}, \quad x = \left(\frac{a}{a_0}\right)^2 \tag{4.40}$$

In the case of the singular and asymptotically singular solutions $0 \leqslant x \leqslant x_+$ ($x = 0$ being the singularity: $a = N = 0$) and as x decreases from x_+ to zero, t and b as functions of a are expressed as

$$t = \frac{2a_0}{|\beta|\sqrt{x_+ - x_-}} F(k,\xi),$$
$$b = \frac{\beta/|\beta|}{\sqrt{x_+ - x_-}} [(x_0 - x_-)F(k,\xi) - (x_+ - x_-)E(k,\xi)], \tag{4.41}$$

up to a time and x^3 translation, where now

$$k = \left(\frac{x_+}{x_+ - x_-}\right)^{1/2}, \quad \xi = \left(\frac{x_+ - x}{x_+}\right)^{1/2}, \quad x = \left(\frac{a}{a_0}\right)^2. \tag{4.42}$$

ACKNOWLEDGMENTS

One of us (A.A.) is grateful to the Department of Physics, University of Sussex, for its hospitality during summer 1978 when part of this work was done.

[1] A. Aurilia and D. Christodoulou "Theory of strings and membranes in an external field. I. General formulation," Imperial College preprint, 77-78/28.
[2] S. Mandelstam, Phys. Rep. **13**, 260 (1974); J. Scherk, Rev. Mod. Phys. **47**, 123 (1975).
[3] M. Kalb and P. Ramond, Phys. Rev. D **9**, 2273 (1974); F. Lund and T. Regge, Phys. Rev. D **14**, 1524 (1976); A. Aurilia and D. Christodoulou, Phys. Lett. **71B**, 90 (1977).
[4] H.B. Nielsen and P. Olesen, Nucl. Phys. **B61**, 45 (1973).
[5] D. Saint-James, G. Sarma, and E.J. Thomas, *Type II Superconductivity* (Pergamon, New York, 1969).
[6] Y. Nambu, Phys. Rev. D **10**, 4262 (1974); A. Aurilia, Nucl. Phys. **B92**, 241 (1975); A. Aurilia and F. Legovini, Phys. Lett. **67B**, 299 (1977).
[7] A. Aurilia, D. Christodoulou, and F. Legovini, Phys. Lett. **73B**, 429 (1978); A. Aurilia and D. Christodoulou, Phys. Lett. **78B**, 589 (1978).

A Lie group framework for composite particles and mass spectrum

Daniel Beau

Physique Mathématique, Faculté des Sciences Mirande, 21000–Dijon, France

Salah Horchani

Département de Mathématiques, Faculté des Sciences de Tunis, Campus Belvédère, Tunis, Tunisia
(Received 1 August 1978)

We propose a concept of internal structure and a relativistically covariant method of unifying the external and internal structures, leading to a dynamical Lie algebra without superfluous generators. In this framework we study in more detail a Lie algebra unifying external space with an internal 3-space, and several representations which describe models of composite particles and give rise to various mass formulas capable of describing the hadron spectrum; we make use of both unitary irreducible global representations and partially integrable, Schur-irreducible, symmetric local representations.

I. INTRODUCTION

The ever-increasing number of particles lead physicists to build models in order to organize the wealth of data in the resonance region for high-energy phenomena.

The success of the quark model, the importance of the harmonic oscillator in nuclear physics, and the fact that SU(3) is a symmetry group of the same, suggested, since the quarks came out, that we look for a (nonrelativistic) hadron model based on the harmonic oscillator.[1] Various relativistic generalizations have been made since then.[2] One of the difficulties in these approaches is that the very concept of the "relativistic" harmonic oscillator is not yet been perfectly defined, and one of the drawbacks is that the relativistic covariance is not always satisfied. The most important (formal) contribution in this respect is that of Feynman *et al.*,[2] which is based on Greenberg's nonrelativistic harmonic oscillator symmetric quark model. Nevertheless, several criticisms can be leveled at their analysis, the most important of which (cf. also Ref. 3) bear on the ambiguity of definition and the spectrality of the intervening operators on the one hand, and on the fact that they disregard the timelike excitations after having postulated a relativistic treatment at the outset.

In this paper, a partial summary of which has been published in Ref. 4, we propose, within the framework of Lie algebras, a relativistic covariant internal formalism and unification method leading to various hadron mass formulas, both in irreducible unitary global representations and in partially integrable, Schur-irreducible, symmetric local representations. The attitude which we have taken here is that of reconsidering the very concept of "internal structure" rather than the symmetries of this structure and of giving the "mass-observable" priority over the "internal observables." In our opinion, the internal symmetries, which manifest themselves in the various interactions of elementary particles, can be considered, at least in a first approximation, as adding constraints to the external degrees of freedom thereby lifting the possible degeneracy of energy. For this last function, at least for the strong mass-differences a perturbative method is difficult to justify. Moreover, there is nothing to prevent the mass from depending on parameters which need not be conserved in the interaction observed at present, i.e., on supplementary quantum numbers that do not appear in the present characterization of the particles, as the quantum numbers have so far been considered chiefly within the framework of conservation laws. All of this is justification for not subjecting our model to any constraint of the unitary symmetry type at the outset. We shall obtain our mass-operator by substituting for the usually considered monoparticular states (responsible for the continuity of the spectrum) multiparticular states with degrees of internal freedom. In this composite particles model the internal structure is described by the Heisenberg algebra $\mathfrak{h}_n$ and the external structure by the Poincaré algebra $\mathfrak{p}$. We unify them by a Lie subalgebra $\mathfrak{g}$, containing $\mathfrak{p}$, of $\mathcal{U}(\mathfrak{h}_n) \otimes \mathcal{U}(\mathfrak{p})$, which is isomorphic to $\mathcal{U}(\mathfrak{h}_n \oplus \mathfrak{p})$ (Ref. 5, corollary 2.2.12); $\mathcal{U}$ denotes the enveloping algebra. We also adopt the following hypothesis: The generators of $\mathfrak{g}$, other than those of the Poincaré algebra, are relativistic covariant; more precisely, they commute with the translations of $\mathfrak{p}$ and constitute a basis of a real finite-dimensional representation (direct sum of irreducible representations) of $\mathfrak{sl}(2,C)$.

The hypothesis that the (relativistic) internal structure is described by the Lie algebra of the (nonrelativistic) commutation relations might, at first sight, seem ambiguous. Now it often happens that nonrelativistic concepts are used in a relativistic context as is the case, for instance, of the parton model in the infinite momentum frame (cf. for example Ref. 6). Moreover, to our knowledge, no experiment mentions any observable difference between the relativistic and nonrelativistic intrinsic internal structures. In addition the same hypothesis, with the Poincaré group replaced by the Galilée group could, in our opinion, be also considered as a nonrelativistic approximation.

The hadrons are interpreted as the excited levels of these composite particles, whose interpretation shows how energy (mass) is created on the basis of the internal dynamics thus defined. The fact that particles can be interpreted in this framework as the excited states of a system composed of harmonically interacting "canonically conjugate" components relates our model to that of Feynman *et al.*[2]

 0022-2488/79/081700-12$01.00

The internal degrees of freedom result therefore, in our model, from the Heisenberg Lie algebra $\mathfrak{h}_n$ and its enveloping algebra. Several reasons suggest that this choice is natural and reasonable, among which we may quote:

(a) $\mathfrak{h}_n$ is at the basis of the mathematical formulation of the nonrelativistic quantum phenomena, being the canonical algebra of observables in standard quantum mechanics.

(b) the Gel'fand–Kirillov conjecture,[7] that if K is a commutative field of zero characteristic, $\mathfrak{g}$ an algebraic Lie algebra on K, $\mathfrak{h}_n$ the Heisenberg algebra on K of generators $(q_1,\ldots,q_n;p_1,\ldots,p_n;1)$, $D_{n,k}(K)$ the field generated on K by the generators $q_1,\ldots,q_n;p_1,\ldots,p_n;z_1,\ldots,z_k$, where all the commutators other than those of $\mathfrak{h}_n$ are zero, then there exists n and k such that the enveloping field of $\mathfrak{g}$ is isomorphic to $D_{n,k}(K)$. This conjecture has been proved in the most important cases, including the complex solvable and semisimple Lie algebras and is therefore true, e.g., for $\mathfrak{sl}(p)$, etc.

(c) $\mathfrak{h}_n$ appears naturally in the structure of the extended Galilée group and the "passive" Galilée and Poincaré groups.[8]

2. DEFINITION AND STRUCTURAL PROPERTIES OF THE GROUPS G_n

Let $(M_{\mu\nu},P_\sigma)$ be the canonical basis of $\mathfrak{p} = R^4 \oplus\!\!\!\!+\ \mathfrak{sl}(2,C)$ (where $\oplus\!\!\!\!+$ denotes the semidirect sum of Lie algebras, the semidirect product of the groups will be denoted by a point, and the Greek indices vary from 1 to 4) and $\mathfrak{g}_n$ be the subalgebra of $\mathcal{U}(\mathfrak{h}_n)\otimes\mathcal{U}(\mathfrak{p})$ generated by

$$L_{\mu\nu} = 1\otimes M_{\mu\nu},\quad T_\mu = 1\otimes P_\mu,\quad Q_{i\mu} = q_i\otimes P_\mu,$$
$$A_{i\mu} = p_i\otimes P_\mu,\quad C_{\mu\nu} = 1\otimes P_\mu P_\nu.$$

So $\mathfrak{g}_n$ is a Lie algebra of dimension $8n+20$ whose (nonzero) commutation relations are given by

$$\begin{aligned}
&[L_{\mu\nu},L_{\rho\sigma}] = -g_{\mu\rho}L_{\nu\sigma} - g_{\nu\sigma}L_{\mu\rho} + g_{\mu\sigma}L_{\nu\rho} + g_{\nu\rho}L_{\mu\sigma},\\
&[L_{\mu\nu},X_\rho] = g_{\nu\rho}X_\mu - g_{\mu\rho}X_\nu,\\
&[L_{\mu\nu},C_{\rho\sigma}] = -g_{\mu\rho}C_{\nu\sigma} + g_{\nu\sigma}C_{\mu\rho} - g_{\mu\sigma}C_{\nu\rho} + g_{\nu\rho}C_{\mu\sigma},\\
&[A_{i\mu},Q_{j\nu}] = \delta_{ij}C_{\mu\nu},
\end{aligned}\tag{1}$$

where X_ρ denotes $T_\rho, A_{i\rho}$, or $Q_{i\rho}$ (for fixed i), and $g_{\alpha\beta}$ is the usual metric tensor ($g_{ij} = \delta_{ij}$ and $g_{4i} = 0$, for $1\leqslant i,j\leqslant 3$, $g_{44} = -1$).

Let $\mathfrak{n}_n$ (resp. R^4) be the subalgebra of $\mathfrak{g}_n$ generated by $(A_{i\mu},Q_{j\nu},C_{\rho\sigma})$ [resp. (T_μ)]. If we denote by $D(j,j')$ the irreducible representation of dimension $(2j+1)(2j'+1)$ of $\mathfrak{sl}(2,C)$, we have the following proposition.

Proposition 2.1: $\mathfrak{g}_n$ is the semidirect sum of $\mathfrak{sl}(2,C)$ by the nilpotent ideal $R^4\oplus\mathfrak{n}_n$ relatively to the representation

$$\left\{\bigoplus_{\alpha=1}^{2n+1} D_\alpha(\tfrac12,\tfrac12)\right\}\oplus D(1,1)\oplus D(0,0),$$

where $D_\alpha(\frac12,\frac12) = D(\frac12,\frac12)$ for all α.

Proof: Let $R_{\underline{c}}^{10}$ be the subalgebra of $\mathfrak{n}_n$ generated by $(C_{\mu\nu})$; it is the center of $\mathfrak{n}_n$ and the quotient algebra $\mathfrak{n}_n/R_{\underline{c}}^{10}$ is isomorphic to R^{8n}.

Then $\mathfrak{n}_n$ is nilpotent as a central extension of commutative Lie algebras. The semidirect sum,

$\mathfrak{g}_n = (R^4\oplus\mathfrak{n}_n)\oplus\!\!\!\!+\ \mathfrak{sl}(2,C)$, is defined by $\{\bigoplus_{\alpha=1}^{2n+1} D_\alpha(\frac12,\frac12)\}\oplus\pi$, where π, the representation of $\mathfrak{sl}(2,C)$ acting in $R_{\underline{c}}^{10}$, is easily identified with the symmetric component of the representation $D(\frac12,\frac12)\oplus D(\frac12,\frac12)$ of $\mathfrak{sl}(2,C)$. Then it follows that $\pi = D(1,1)\oplus D(0,0)$.

Q.E.D.

Now we shall denote by $\underline{x}$ [resp. $\tilde{x}$, $\vec{x}$, $\underline{x}$, $(\underline{x}_i)_{1<i<n}$, or $\times_{i=1}^n \underline{x}_i, A$] a 4-vector [resp. the associated 2-vector, the associated 3-vector, the generic element of R^{10}, R^{4n}, $\mathfrak{sl}(2,C)$], $\theta^{\mu\nu}$ the function equal to $\frac12$ if $\mu=\nu$ and to 1 if not, $H^2(R^n,R^m)$ the second cohomology group of R^n with coefficients in R^m and $h = \{(\underline{a}_i)(\underline{q}_i)\}$ [resp., $g = \{t,\underline{c},(\underline{a}_i),(\underline{q}_i),A\}$] the generic element of the group R^{8n} (resp. of the connected and simply connected group G_n of Lie algebra $\mathfrak{g}_n$.

Proposition 2.2: The group law of G_n is given by:

$$\begin{aligned}
gg' = \{&\underline{t}+D(\tfrac12,\tfrac12)(A)\underline{t}',\underline{c}+[D(1,1)\oplus D(0,0)](A)\underline{c}'\\
&+\underset{\sim}{\beta}(h,D(\tfrac12,\tfrac12)(A)h'),(\underline{a}_i+D(\tfrac12,\tfrac12)(A)\underline{a}_i'),\\
&(\underline{q}_i+D(\tfrac12,\tfrac12)(A)\underline{q}_i'),AA'\},
\end{aligned}$$

where $\underset{\sim}{\beta}\in H^2(R^{8n},R^{10})$ which is defined by

$$\underset{\sim}{\beta}(h,h')^{\mu\nu} = \theta^{\mu\nu}\delta^{ij}\{(\underline{a}_i\otimes\underline{q}'_j)^{\mu\nu} + (\underline{a}_i\otimes\underline{q}'_j)^{\nu\mu}\}$$

and

$$D(\tfrac12,\tfrac12)(A)h = \{(D(\tfrac12,\tfrac12)(A)\underline{a}_i),(D(\tfrac12,\tfrac12)(A)\underline{q}_i)\}.$$

Proof: Let $K = R^{10}\times_{\underset{\sim}{\beta}} R^{8n}$ be the connected and simply connected Lie group, central extension of R^{8n} by R^{10} defined by the factor $\underset{\sim}{\beta}$. If $\{\underline{c},(\underline{a}_i),(\underline{q}_j)\}$ denotes its generic element, its group law is given by

$$\begin{aligned}
&\{\underline{c},(\underline{a}_i),(\underline{q}_j)\}\{\underline{c}',(\underline{a}'_i),(\underline{q}'_j)\}\\
&\quad = \{\underline{c}+\underline{c}'+\underset{\sim}{\beta}(h,h'),(\underline{a}_i+\underline{a}'_i),(\underline{q}_j+\underline{q}'_j)\}.
\end{aligned}$$

Let us determine the Lie algebra $\mathfrak{k}$ of K; for that we consider the following one-parameter subgroups of K:

$$k:t\to\{t\underline{c},\underline{0},\underline{0}\},\quad k':t\to\{\underline{0},(t\underline{a}_i),\underline{0}\},\quad k'':t\to\{\underline{0},\underline{0}(t\underline{q}_i)\}.$$

The vector space subjacent to $\mathfrak{k}$ is generated by the $f'(t)|_{t=0}$, where $f\in(k,k',k'')$. It is isomorphic to $R^{10}\times R^{8n}$. The only nonzero commutators of $\mathfrak{k}$ are those relative to the products $\{\underline{0},(\underline{a}_i),\underline{0}\}\ \{\underline{0},\underline{0},(\underline{q}_j)\}$; more precisely, they are given by

$$[\{\underline{0},(\underline{a}_i),\underline{0}\},\{\underline{0},\underline{0},(\underline{q}_j)\}] = \gamma'(t)|_{t=0}$$

where γ is the curve;

$$\gamma:t\to\{\underline{0},(t^{1/2}\underline{a}_i),\underline{0}\}\{\underline{0},\underline{0},(t^{1/2}\underline{q}_j)\}\{\underline{0},(t^{1/2}\underline{a}_i),\underline{0}\}^{-1}\{\underline{0},\underline{0},(t^{1/2}\underline{q}_j)\}^{-1}.$$

We obtain

$$[\{\underline{0},(\underline{a}_i),\underline{0}\},\{\underline{0},\underline{0}(\underline{q}_j)\}] = \{\underline{\beta}(\{(\underline{a}_i),\underline{0}\},\{\underline{0},(\underline{q}_j)\}),\underline{0},\underline{0}\}.$$

Suppose $A'_{i\mu}$ (resp. $Q'_{i\mu}, C'_{\mu\nu}$) is the vector $\{\underline{0},(\underline{a}_j),\underline{0}\}$ (resp. $\{\underline{0},\underline{0},(\underline{q}_j)\}$, $\{\underline{c},\underline{0},\underline{0}\}$), all the components of which are zero save for $(a_i)^\mu$ [resp. $(q_j)_\mu, c^{\mu\nu}$] which is equal to 1. The previous relation then gives $[A'_{i\mu},Q'_{j\nu}] = \delta_{ij}C'_{\mu\nu}$. This implies that $\mathfrak{k}$ is isomorphic to $\mathfrak{n}_n$, and the result is a direct consequence of Proposition 2.1.

Remarks 2.1: (a) If we denote by C (resp. T,G) the symmetric matrix with coefficients $C_{\mu\nu}$ (resp. $T_\mu T_\nu, g_{\mu\nu}$), the ele-

ments Tr $[GC]^k$ ($k = 1,2,3,4$) and Tr$[(GT)(GC)^l]$ ($l = 0,1,2,3$), where Tr denotes the trace, are in the center of $\mathcal{U}$ $[(R^4 \oplus R^{10}) \oplus \mathfrak{sl}(2,C)]$.

(b) Given one of the algebras $\mathfrak{g}_n$ and another Lie algebra $\mathfrak{l}_n$ of dimension n, it is possible to construct in a natural way an algebra $U(\mathfrak{g}_n,\mathfrak{l}_n)$ unifying[9] then in the following manner: If $[X_i,X_j] = C_{ij}{}^k X_k$ are the structural relations of $\mathfrak{l}_n$, then $U(\mathfrak{g}_n,\mathfrak{l}_n)$ is generated by $(\mathfrak{g}_n,\mathfrak{l}_n)$ with their respective structural relations to which we adjoin

$$[X_i,Q_{j\mu}] = C_{ij}{}^k Q_{k\mu}, \quad [X_i,A_{j\mu}] = C_{ij}{}^k A_{k\mu}.$$

Thus defined, $U(\mathfrak{g}_n,\mathfrak{l}_n)$ admits the decompositions

$$U(\mathfrak{g}_n,\mathfrak{l}_n) = \mathfrak{g}_n \oplus \mathfrak{l}_n = (R^4 \oplus \mathfrak{n}_n) \oplus (\mathfrak{sl}(2,C) \oplus \mathfrak{l}_n)$$
$$= \{R^4 \oplus (\mathfrak{n}_n \oplus \mathfrak{l}_n)\} \oplus \mathfrak{sl}(2,C),$$

where $\mathfrak{n}_n \oplus \mathfrak{l}_n$ is defined by the direct sum of ten trivial representations and eight representations equivalent to the adjoint representation of $\mathfrak{l}_n$.

3. DEFORMATIONS OF SOME SUBALGEBRAS OF $\mathfrak{g}_n$

In order to bring to light other possibly interesting models and the "internal symmetries" content of the groups G_n, we are going, in this article, to consider some deformations of certain subalgebras of $\mathfrak{g}_n$ (deformations not necessarily extensible to $\mathfrak{g}_n$) which present a potential physical interest corroborated by the new particles recently discovered which lead, among other things, to higher internal unitary symmetries.

Proposition 3.1: $R_c^{10} \oplus \mathfrak{sl}(2,C)$ can be deformed in the first order into $\mathfrak{u}(3,1)$, $\mathfrak{gl}(4,R)$, and $\mathfrak{w}_9$, where $\mathfrak{w}_9$ is the Weyl algebra associated with the representation $D(1,1)$ of $\mathfrak{sl}(2,C)$.

Proof: The bracket $[\ ,\ ]_t$ of such a deformation is necessarily written

$$[X_1,X_2]_t = [X_1,X_2] + t\psi(X_1,X_2), \quad \forall\ X_1,X_2 \in \mathfrak{g},$$

where $\mathfrak{g} = R_c^{10} \oplus \mathfrak{sl}(2,C)$ and ψ is an element of $\mathscr{H}^2(\mathfrak{g},\mathfrak{g})$. A line of argument analogous to that of Ref. 10 shows that it is sufficient to restrict ourselves to $\psi \in \mathscr{H}^2(R_c^{10},\mathfrak{g})^{\mathfrak{g}}$, the second cohomology group relative to $\mathfrak{sl}(2,C)$. Thus the conditions of integrability come down to:

$$\sum_{P(X_1,X_2,X_3)} \psi(\psi(X_1,X_2),X_3) = 0, \quad \forall\ X_1,X_2,X_3 \in R_c^{10},$$

where $P(X_1,X_2,X_3)$ = circular permutation on (X_1,X_2,X_3).

(a) Let $\psi \in \mathscr{H}^2(R_c^{10},\mathfrak{g})^{\mathfrak{g}}$ be defined by

$$\psi(C_{\mu\nu},C_{\rho\sigma}) = -g_{\mu\rho}L_{\nu\sigma} - g_{\nu\sigma}L_{\mu\rho} - g_{\mu\sigma}L_{\nu\rho} - g_{\nu\rho}L_{\mu\sigma}.$$

This cocycle verifies the conditions of integrability and defines thus a Lie algebra law.

Denoting $|t|^{-1/2} C_{\mu\nu}$ by $N_{\mu\nu}$ and according as $t > 0$ or $t < 0$, we recognize the commutation relations of $\mathfrak{u}(3,1)$ or $\mathfrak{gl}(4,R)$, respectively.

(b) Let $\psi \in \mathscr{H}^2(R_c^{10},\mathfrak{g})^{\mathfrak{g}}$ be defined by

$$\psi(D,F_i) = F_i, \quad \psi(F_i,F_j) = 0,$$

where $\{D,(F_i)\ 1 \leqslant i \leqslant 9\}$ is a basis of R_c^{10} such that D generates $D(0,0)$ and (F_i), $1 \leqslant i \leqslant 9$, generates $D(1,1)$.

Then ψ defines a deformation of $\mathfrak{g}$ isomorphic to the Lie algebra of the group $W_9 = (R^9 \cdot R_+^\times) \cdot \mathfrak{sl}(2,C)$; $R_+^\times$ being the multiplicative group of positive reals, the semidirect products being defined, respectively, by the representation $D(1,1)$ of $\mathfrak{sl}(2,C)$, the trivial homomorphism of $\mathfrak{sl}(2,C)$ in Aut $R_+^\times$ $(= R^\times)$ and the homomorphism of $R_+^\times$ in Aut $(R^9) = \mathfrak{gl}(9,R)$ which makes the homothetic transformation of ratio λ correspond to λ.

Corollary 3.1:

1. $R^9 \oplus \mathfrak{sl}(2,C)$ [resp. $R^5 \oplus \mathfrak{su}(2)$, defined by the representation of weight 2] deforms in these various deformations into $\mathfrak{su}(3,1)$, $\mathfrak{sl}(4,R)$, and itself [resp. $\mathfrak{su}(3)$, $\mathfrak{sl}(3,R)$, and itself].

2. All the deformations considered above contract into their initial Lie algebra.

Remarks 3.1:

(a) As $R^9 \oplus \mathfrak{sl}(2,C)$ is a stable special (on the field of the reals) affine Lie algebra,[11] $\mathrm{Der}(R^9 \oplus \mathfrak{sl}(2,C)) = \mathfrak{w}_9$.

(b) The invariants Tr$[(GC)^k]$ of $\mathfrak{g}_n$ are the contraction of the invariants of $\mathfrak{u}(3,1)$ and generate the center of $\mathcal{U}(R_c^{10} \oplus \mathfrak{sl}(2,C))$.

(c) $\mathfrak{u}(3,1)$ and $\mathfrak{gl}(4,R)$, being reductive with one-dimensional center, are rigid.

Proposition 3.2: Let $R^{18} \oplus \mathfrak{sl}(2,C)$ be the subalgebra of $\mathfrak{g}_n$ generated by $(C_{\mu\nu},L_{\rho\sigma},X_\alpha,Y_\beta)$, where (X_α) and (Y_β) are two 4-vectors of $\mathfrak{g}_n$ such that $[X_\alpha,Y_\beta] = 0$. So, $R^{18} \oplus \mathfrak{sl}(2,C)$ deforms into $R^8 \oplus \mathfrak{u}(3,1)$, $\mathfrak{su}(p,q)$ $(p+q=5, pq \neq 0)$, $R^8 \oplus \mathfrak{gl}(4,R)$ and $\mathfrak{sl}(5,R)$; where $R^8 \oplus \mathfrak{u}(3,1)$ is generated by $\mathfrak{u}(3,1)$ and the complex translations and $R^8 \oplus \mathfrak{gl}(4,R)$ is defined by the direct sum of the basic representation of $\mathfrak{gl}(4,R)$ and of its coadjoint.

Proof: (a) Let $\mathfrak{g}_t$ be the second order deformation of $\mathfrak{g} = R^{18} \oplus \mathfrak{sl}(2,C)$, the bracket of which is given by

$$[\ ,\ ]_t = [\ ,\ ] + t\psi_1 + t^2\psi_2, \tag{2}$$

where ψ_1 and $\psi_2 \in \mathscr{C}^2(R^{18},\mathfrak{g})^{\mathfrak{g}}$, the space of 2-cochains relative to $\mathfrak{sl}(2,C)$, such that;

$$\psi_1(C_{\mu\nu},C_{\rho\sigma}) = 0, \quad \psi_1(C_{\mu\nu},X_\sigma) = -g_{\nu\sigma}Y_\mu - g_{\mu\sigma}Y_\nu,$$
$$\psi_1(C_{\mu\nu},Y_\sigma) = g_{\nu\sigma}X_\mu + g_{\mu\sigma}X_\nu,$$
$$\psi_2(C_{\mu\nu},C_{\rho\sigma}) = -g_{\mu\rho}L_{\nu\sigma} - g_{\nu\sigma}L_{\mu\rho} - g_{\mu\sigma}L_{\nu\rho} - g_{\nu\rho}L_{\mu\sigma}.$$

Denoting $tL_{\mu\nu}$ by $L'_{\mu\nu}$, we recognize the commutation relations of the Lie algebra $R^8 \oplus \mathrm{U}(3,1)$ of the group $C^4.\mathrm{U}(3,1)$.

(b) Let $\mathfrak{g}_{t,\theta}$ be the first order deformation of $\mathfrak{g}_t = R^8 \oplus \mathfrak{u}(3,1)$ above, the bracket of which is defined by

$$[\ ,\ ]_{t,\theta} = [\ ,\ ]_t + \theta\psi, \tag{3}$$

where ψ is the 2-cochain $\in \mathscr{C}^2(R^8,\mathfrak{g}_t)^{\mathfrak{g}_t}$ defined by

$$\psi(X_\nu,X_\sigma) = tL'_{\nu\sigma}, \quad \psi(X_\nu,Y_\sigma) = tC_{\nu\sigma} + tg_{\nu\sigma}g^{\alpha\beta}C_{\alpha\beta},$$
$$\psi(Y_\nu,Y_\sigma) = tL'_{\nu\sigma}.$$

If we denote

$$L'_{5\nu} = |\theta|^{-1/2} X_\nu = -L'_{\nu 5}, \quad C_{5\nu} = |\theta|^{-1/2} Y_\nu = C_{\nu 5},$$
$$g_{55} = \text{sign}(-\theta), \quad C_{55} = -g_{55} g^{\alpha\beta} C_{\alpha\beta},$$

the commutation relations of $\mathfrak{g}_{t,\theta}$ are those of $\mathfrak{su}(p,q)$ $(p+q=5, pq\neq 0)$, so $\mathfrak{g}_{t,\theta}$ is isomorphic to $\mathfrak{su}(3,2)$ [resp. $\mathfrak{su}(4,1)$] if $\theta > 0$ (resp. $\theta < 0$)

(c) Let $\mathfrak{g}_t$ be the second order deformation of $\mathfrak{g} = R^{18} \oplus \mathfrak{sl}(2,C)$, the bracket of which is given by (2) with ψ_1 and $\psi_2 \in \mathscr{C}^2(R^{18}, \mathfrak{g})^{\mathfrak{g}}$ defined by

$$\psi_1(C_{\mu\nu}, C_{\rho\sigma}) = 0, \quad \psi_1(C_{\mu\nu}, X_\sigma) = -g_{\nu\sigma} X_\mu - g_{\mu\sigma} X_\nu,$$
$$\psi_1(C_{\mu\nu}, Y_\sigma) = g_{\nu\sigma} Y_\mu + g_{\mu\sigma} Y_\nu,$$
$$\psi_2(C_{\mu\nu}, C_{\rho\sigma}) = g_{\mu\rho} L_{\nu\sigma} + g_{\nu\sigma} L_{\mu\rho} + g_{\mu\sigma} L_{\nu\rho} + g_{\nu\rho} L_{\mu\sigma}.$$

Denoting $tL_{\mu\nu}$ by $L'_{\mu\nu}$, we easily identify the commutation relations of $R^8 \oplus \mathfrak{gl}(4,R)$ where the semidirect sum is defined by the direct sum of the basic representation of $\mathfrak{gl}(4,R)$ and of its coadjoint [R^4 being endowed with the metric tensor $(tg_{\mu\nu})$].

(d) Let $\mathfrak{g}_{t,\theta}$ be the first order deformation of $\mathfrak{g}_t = R^8 \oplus \mathfrak{gl}(4,R)$ above, written in the basis $\{L'_{\mu\nu}; C_{\mu\nu}; U_\sigma = X_\sigma + Y_\sigma; V_\sigma = X_\sigma - Y_\sigma\}$, the bracket being defined by (3) where ψ is the 2-cochain $\in \mathscr{C}^2(R^8, \mathfrak{g}_t)^{\mathfrak{g}_t}$ defined by

$$\psi(U_\nu, U_\sigma) = tL'_{\nu\sigma}, \quad \psi(U_\nu, V_\sigma) = tC_{\nu\sigma} + tg_{\nu\sigma} g^{\alpha\beta} C_{\alpha\beta},$$
$$\psi(V_\nu, V_\sigma) = -tL'_{\nu\sigma}.$$

If we denote

$$L'_{5\nu} = |\theta|^{-1/2} U_\nu = -L'_{\nu 5}, \quad C_{5\nu} = |\theta|^{-1/2} V_\nu = C_{\nu 5},$$
$$g_{55} = \text{sign}(-\theta), \quad C_{55} = -g_{55} g^{\alpha\beta} C_{\alpha\beta},$$

the commutation relations of $\mathfrak{g}_{t,\theta}$ are those of $\mathfrak{sl}(5,R)$.

Proposition 3.3: Let $\mathfrak{l}_3 = \mathfrak{so}(3)$ or $\mathfrak{so}(2,1)$, the corresponding metric being denoted by (γ_{ij}), and $R^{12} \oplus (\mathfrak{sl}(2,C) \oplus \mathfrak{l}_3)$ be the subalgebra of $U(\mathfrak{g}_3, \mathfrak{l}_3)$ generated by $(\mathfrak{sl}(2,C), \mathfrak{l}_3)$ and $(A_{i\mu})$. Then $R^{12} \oplus (\mathfrak{sl}(2,C) \oplus \mathfrak{l}_3)$ deforms into $\mathfrak{so}(p,q)$ $(p+q=7, pq\neq 0)$, the metric being $(g_{\nu\sigma}) \oplus (\gamma_{ij})$.

Proof: Let $\mathfrak{g} = R^{12} \oplus (\mathfrak{sl}(2,C) \oplus \mathfrak{l}_3)$ and $\mathfrak{g}_t$ be the "second order" deformation of $\mathfrak{g}$ with bracket

$$[\ ,\ ]_t = [\ ,\ ] + t^2 \psi,$$

the 2-cochain $\psi \in \mathscr{C}^2(R^{12}, \mathfrak{g})^{\mathfrak{g}}$ being defined by

$$\psi(A_{i\mu}, A_{j\rho}) = \gamma_{ij} L_{\rho\mu} - g_{\rho\mu} M_{ij},$$

where M_{ij} are the generators of $\mathfrak{l}_3$. If we write

$$L_{4+i,4+j} = M_{ij}, \quad L_{\mu,4+i} = (1/t) A_{i\mu} = -L_{4+i,\mu},$$
$$g_{4+i,4+j} = \gamma_{ij},$$

we find that $(L_{ab};\ 1 \leqslant a,b \leqslant 7)$ verify the commutation relations of $\mathfrak{so}(p,q)$ $(p+q=7, pq\neq 0)$ with metric (g_{ab}).

Remarks 3.2: (a) All the deformations of the two preceding propositions evidently contract into their initial Lie algebra.

(b) One of the advantages of the deformations of the preceding proposition comes from the appearance of $\mathfrak{so}(4,3)$ which contains $\mathfrak{g}_{2,2}$ [the (only) noncompact real form of the complex minimal-dimensional exceptional semisimple Lie algebra]. This Lie algebra appeared in the search for a nontrivial bond between external and internal symmetries.

More explicitly,[9] $\mathfrak{so}(4,3) = \mathrm{U}(\mathfrak{p}, \mathfrak{g}_{2,2})$ with intersection on $\mathfrak{h}_1$, and any other unification of $\mathfrak{p}$ and $\mathfrak{g}_{2,2}$ is trivial.

(c) Let $\mathfrak{s}_p$ $(1 \leqslant p < n)$ (resp. $\mathfrak{s}_n$) be the subalgebra of $\mathfrak{g}_n$ of basis $\{L_{\mu\nu};\ C_{\rho\sigma};\ A_{p\alpha};\ Q_{p+1,\beta}\}$ (resp. $\{L_{\mu\nu};\ C_{\rho\sigma};\ A_{n\alpha};\ Q_{1\beta}\}$. Then $\mathfrak{g}_n = U(\mathfrak{p}, \mathfrak{s}_1, \ldots, \mathfrak{s}_n)$ and (for all $p = 1, \ldots, n$) $\mathfrak{s}_p$ deforms according to Proposition 3.2.

4. IRREDUCIBLE UNITARY REPRESENTATIONS OF THE GROUP G_3

Let R_a^{4n} (resp. R_q^{4n}) be the subgroup of G_n generated by $(A_{i\mu})$ [resp. $(Q_{i\mu})$]. Then G_n admits the decomposition

$$G_n = R^4 \cdot \{(R_c^{10} \times R_a^{4n}) \cdot (R_q^{4n} \cdot \mathrm{SL}(2,C))\}.$$

Consequently, the most natural method of determining its strongly continuous unitary irreducible representations (UIR) is the method of induced representations[12] (in stages), provided that it turns out to have the required properties. In what follows we shall restrict ourselves to G_3; the treatment of G_n, for any n would be similar.

Proposition 4.1: Let $G_3 = H \cdot K$, where H (resp. K) is the group R^4 [resp. $N_3 \cdot \mathrm{SL}(2,C)$], $\hat{H}$ is the dual of H, $\underline{p}$ is the generic element of $\hat{H}$. Then:

1. The orbits of $\hat{H}$, relatively to the action of K, are the supersurfaces $\Omega_+^{m^2}$ $(\underline{p}^2 = -m^2, p^4 > 0)$; $\Omega_-^{m^2}$ $(p^2 = -m^2, p^4 < 0)$; Ω^{-m^2} $(p^2 = m^2)$; Ω_+^0 $(p^2 = 0, p^4 > 0)$; Ω_-^0 $(\underline{p}^2 = 0, p^4 < 0)$, and Ω_0^0 $(\underline{p} = 0)$ with the respective stabilizers $N_3 \cdot \mathrm{SU}(2)$, $N_3 \cdot \mathrm{SU}(2)$, $N_3 \cdot \mathrm{SL}(2,R)$, $N_3 \cdot \widetilde{E}_2$, $N_3 \cdot \widetilde{E}_2$, and $N_3 \cdot \mathrm{SL}(2,C)$, where m is a strictly positive real number and $\widetilde{E}_2$ is the two-fold covering of the Euclidean group E_2 in two dimensions.

2. The decomposition $H \cdot K$ of the group G_3 is a regular semidirect product; the subgroups H and K are closed in G_3 and G_3 is a separable locally compact group.

Proof: The action of K on $\hat{H}$ is defined by

$$\langle k[\hat{h}], h \rangle = \langle \hat{h}, k^{-1} h k \rangle, \ \forall\, h \in H,$$

where k (resp. $\hat{h}$) is an element of K (resp. $\hat{H}$) and $\langle\ ,\ \rangle$ denotes the action of $\hat{H}$ on H, defined here by

$$\langle \hat{h}, h \rangle = \exp i(\underline{p}\underline{t}), \text{ where } \underline{p} \in \hat{H}, \underline{t} \in H, \underline{p}\underline{t} = g_{\mu\nu} t^\nu p^\mu.$$

If we write $k = \{\underline{0}, \underline{c}, (a_i), (q_i), A\}$, the action of k on $\hat{H}$ reduces to: $\underline{p} \in \hat{H} \to D(\tfrac{1}{2}, \tfrac{1}{2})(A)\underline{p}$; this gives the normal orbits of the Poincaré group.[13]

Remark 4.1: In view of the physical applications we wish to draw from our model, we shall confine ourselves in this article to the determination of the UIR of G_3, for which the Poincaré mass-operator $-g^{\alpha\beta} P_\alpha P_\beta$ is strictly positive (and Poincaré energy is positive, for instance). So it is enough to determine the UIR of the stabilizer $N_3 \cdot \mathrm{SU}(2)$.

Lemma 4.1: Let $G = H \cdot K$, where H (resp. K) is the group $R_c^{10} \times R_a^{12}$ [resp $R_q^{12} \cdot \mathrm{SU}(2)$], $\hat{h} = \{\underline{0}, \underline{d}, (\underline{b}_i), 0, 1\}$ the generic element of $\hat{H} = R_d^{10} \times R_b^{12}$, $\underline{d}^{[\mu\nu]}$ $(\mu \leqslant \nu)$ the components of $\underline{d}$ and DG the matrix with the generic element $d_\lambda^\mu = g_{\lambda\nu} d^{\mu\nu}$, where $d^{\mu\nu} = d^{\nu\mu} = d^{[\mu\nu]}$. Then the action of $k = \{\underline{0}, \underline{0}, \underline{0}, (\underline{q}_i), A\} \in K$ on $\hat{H}$ is given by

$$k(\{\underline{0}, \underline{d}, (\underline{b}_i), \underline{0}, 1\}) = \{\underline{0}, \underline{d}', (\underline{b}'_i), \underline{0}, 1\},$$

where

$$D'G = D(\tfrac{1}{2},\tfrac{1}{2})\ (A)\ DG\ D(\tfrac{1}{2},\tfrac{1}{2})\ (A^{-1}),$$

$$\underline{b}'_i = D(\tfrac{1}{2},\tfrac{1}{2})\ (A)\underline{b}_i + D'G\underline{q}_i.$$

Proof: Let $h = \{\underline{0}, \underline{c}, (\underline{a}_i), \underline{0}, 1\} \in H$. The action of H on $\hat{H}$ is defined by

$$\langle \hat{h}, h \rangle = \exp i\{\delta^{kl}(\underline{b}_k \cdot \underline{a}_l) + \underline{d}\cdot\underline{c}\},$$

where

$$\underline{b}_k \cdot \underline{a}_l = g_{\mu\nu}\underline{b}_k^{\mu}\underline{a}_l^{\nu} \text{ and } \underline{d}\cdot\underline{c} = (g\otimes g)_{[\mu\nu][\rho\sigma]}\underline{d}^{[\mu\nu]}\underline{c}^{[\rho\sigma]},$$

g being the metric tensor of Lorentz and $\{(g\otimes g)_{[\mu\nu][\rho\sigma]}\}$ the Lorentz invariant metric tensor of R_c^{10} obtained by restricting $g\otimes g$ (defined on $R^4\otimes R^4$) to $S(R^4\otimes R^4)$.

The explicit determination of the action of K on $\hat{H}$ is facilitated by the change of basis, $\{C_{\mu\mu}\to C_{\mu\mu};\ C_{\mu\nu}\to 2C_{\mu\nu}$, if $\mu\neq\nu\}$, to which corresponds the factor β_0 (equivalent to the factor β of Proposition 2.2) which is defined by

$$\beta_0^{\mu\nu}(h,h') = (\tfrac{1}{2})\delta^{ij}\{(\underline{a}_i\otimes\underline{q}'_j)^{\mu\nu} + (\underline{a}_i\otimes\underline{q}'_j)^{\nu\mu}\}$$

and by the use of the fact that $D(1,1)\oplus D(0,0)$ is equivalent to $D(\frac{1}{2},\frac{1}{2})\otimes D(\frac{1}{2},\frac{1}{2})$ restricted to $S(R^4\otimes R^4)$. Q.E.D.

Notation:

$U(1)\cdot Z_2$ notes the twofold covering of the semidirect product defined by

$$(\varphi,d)\ (\varphi',d') = (\varphi + d\varphi', dd') \quad \forall\ \varphi\in U(1),$$
$$\forall\ d\in Z_2 = \{\pm 1\};$$

Q_8 denotes the (eight-element) quaternionic group [(σ_l) being the three Pauli matrices]: $Q_8 = \{\pm 1;\ \pm i\sigma_l\ (1\leqslant l\leqslant 3)\}$;

Q_4 is one of the following cyclic subgroups of Q_8,

$$C_4 = \{\pm 1; \pm i\sigma_2\}, \quad C'_4 = \{\pm 1; \pm i\sigma_3\}.$$

Lemma 4.2: Let $G = R_c^{10}\cdot SU(2)$. Then SU(2) operating on R_d^{10}, dual of R_c^{10}, admits as stabilizers the following groups: SU(2) [1]; $U(1)\cdot Z_2$ [1]; U(1) [2]; Q_8 [1]; C_4 [1]; C'_4 [1]; Z_2 [2], where the number in brackets is the number of orbits having the same stabilizer (up to isomorphism).

Proof: According to Lemma 4.1, the action of SU(2) on R_d^{10} is defined by

$$D'G = D(\tfrac{1}{2},\tfrac{1}{2})\ (A)\ DG\ D(\tfrac{1}{2},\tfrac{1}{2})\ (A^{-1}), \quad A\in SU(2). \quad (4)$$

Now DG is written

$$DG = \begin{pmatrix} [D] & -\vec{D} \\ {}^t\vec{D} & -d^{44} \end{pmatrix},$$

where $[D]$ (resp. $\vec{D}$; ${}^t\vec{D}$) is the symmetric 3×3 matrix with generic element d^{ij} [resp. the vector of components (d^{14}, d^{24}, d^{34}); the transpose of $\vec{D}$]. So the problem is reduced to solving the system:

$$[D'] = D^1(A)[D]D^1(A^{-1}) \quad (\text{in } R^6),$$
$$\vec{D}' = D^1(A)\vec{D} \quad (\text{in } R^3),$$
$$d'^{44} = d^{44} \quad (\text{in } R),$$

where $D^1(A)$ is the representation of A in the representation of weight 1 of SU(2).

The action in R^6 being defined by a relation of similitude between real symmetric matrices with $D^1(A)\in SO(3)$, the orbits are characterized by the three possible Jordan forms of such matrices, namely $\Omega(\lambda,\lambda,\lambda)$, $\Omega(\lambda,\lambda,\lambda_1)$, and $\Omega(\lambda_1,\lambda_2,\lambda_3)$. The orbits in R^3 are evidently the spheres S_k of radius $k\geqslant 0$. As for the orbits in R, they are, by definition, reduced to a point $\{\alpha\}$, $\alpha\in R$, and therefore the orbits of R_d^{10} are obtained by considering the Cartesian product $X\times\{\alpha\}$, where X ranges over the set of orbits in R^9 and α ranges over R. So it is sufficient for us to determine the orbits in $R^9 = R^6\times R^3$. First, the orbits in R^9 having a trace reduced to a point on R^6 or R^3 are the following Cartesian products of orbits:

$$\Omega(\lambda,\lambda,\lambda)\times S_k (k>0), \quad \Omega(\lambda,\lambda,\lambda)\times\{0\},$$
$$\Omega(\lambda,\lambda,\lambda_1)\times\{0\}, \quad \Omega(\lambda_1,\lambda_2,\lambda_3)\times\{0\}.$$

The other orbits can be obtained in stages, in a way analogous to that of Ref. 14, as follows:

For each orbit Ω of R^6 there are in $R^6\times R^3$ as many orbits as there are orbits in S_k under the action of the stabilizer of any particular point of Ω. We characterize an orbit of $R^6\times R^3$ by a point $[D]_\Omega$ of Ω and an orbit ω of S_k generated by the stabilizer of $[D]_\Omega$; we write this orbit $[\Omega,\omega]$: It is generated by $([D]_\Omega, \vec{x})$, $\vec{x}\in\omega$.

In this way we find the following orbits in R_d^{10}:

$$\Omega(\lambda,\lambda,\lambda)\times\{0\}\times\{\alpha\},$$
$$\Omega(\lambda,\lambda,\lambda_1)\times\{0\}\times\{\alpha\}, \quad \Omega(\lambda_1,\lambda_2,\lambda_3)\times\{0\}\times\{\alpha\},$$
$$\Omega(\lambda,\lambda,\lambda)\times S_k\times\{\alpha\}, \quad [\Omega(\lambda,\lambda,\lambda_1),\Sigma_k(k)]\times\{\alpha\},$$
$$[\Omega(\lambda,\lambda,\lambda_1),\Sigma_k(0)]\times\{\alpha\}, \quad [\Omega(\lambda,\lambda,\lambda_1),\Sigma_k(x^3)]\times\{\alpha\},$$
$$[\Omega(\lambda_1\lambda_2\lambda_3),\Gamma_k^{(2)}]\times\{\alpha\}, \quad [\Omega(\lambda_1,\lambda_2,\lambda_3),\Gamma_k^{(4)}]\times\{\alpha\},$$

with their respective stabilizers: SU(2), $U(1)\cdot Z_2$, Q_8, U(1), U(1), C_4, Z_2, C'_4, and Z_2.

Here S_k denotes the sphere of radius $k>0$ in R^3; $\Sigma_k(x^3)$ is the union of two small circles of S_k in the planes $\pm x^3$; $0<x^3<k$ and $(\Gamma_k^{(2)}, \Gamma_k^{(4)})$ are the two types of orbits of Q_8 in S_k.

Proposition 4.2: Let us keep the notation of Lemma 4.1. Let $\hat{h}$ be any element of $R_d^{10}\times R_b^{12}$ and $S_{\hat{h}}$ its stablizer relatively to the action of $R_q^{12}\cdot SU(2)$. Then:

1. If $\det DG\neq 0$, $S_{\hat{h}}$ is isomorphic with one of the following groups:

SU(2) [1], $U(1)\cdot Z_2$ [1], U(1) [2], Q_8 [1],
C_4 [1], C'_4 [1], Z_2 [2].

2. The $\det DG = 0$, $S_{\hat{h}}$ is isomorphic with one of the following groups:

$R^3\times(R^9\cdot SU(2))$[1], $R^9\cdot SU(2)$[1],
$R^3\times SU(2)$[1],
$R^3\times R^6\cdot(U(1)\cdot Z_2))$[1], $R^3\times R^3\cdot(U(1)\cdot Z_2))$[1],
$R^6\cdot(U(1)\cdot Z_2)$[1], $R^3\cdot(U(1)\cdot Z_2)$[1],
$R^3\times(U(1)\cdot Z_2)$[1], $R^6\times(R^6\cdot U(1))$[3],
$R^6\cdot U(1)$ [2], $R^6\times U(1)$ [3], $R^3\times(R^6\cdot U(1))$[4],
$R^3\times U(1)$ [5]
$R^3\times(R^3\cdot Q_8)$ [1], $R^3\cdot Q_8$ [1], $R^3\times Q_8$ [1]
$R^6\times(R^3\cdot Q_4)$ [3], $R^6\times Q_4$ [3], $R^3\times(R^3\cdot Q_4)$ [5]
$R^3\times Q_4$ [5], $R_3\cdot Q_4$ [3],
$R^{12}\times Z_2$ [3], $R^9\times Z_2$ [9], $R^6\times Z_2$ [16],
$R^3\times Z_2$ [14],

where the number in brackets is the number of orbits having the same stabilizer (up to isomorphism) and the semidirect products $R^{3q}\cdot L$ [L subgroup of SU(2)] are defined by the restriction of the representation $\oplus_{q=1}^{3} D_q^1$ (where $D_q^1 = D^1$, for all q), from SU(2) to L.

Proof: We first determine the orbits in $R_d^{10}\times R_b^{12}$ under the action of the group $R_q^{12}\cdot$SU(2) and then the respective stablizers. To do so, it is necessary simultaneously to solve the following three linear systems:

$$b_i' = D(\tfrac{1}{2},\tfrac{1}{2})(A)b_i + D'Gq_i, \quad i = 1,2,3. \tag{5}$$

According to the demonstration of Lemma 4.2, we have to distinguish two cases according as det DG is zero or not.

1. det$DG\neq 0$: If we denote by Ω_i $(1\leqslant i\leqslant 9)$ the nine orbits of SU(2) in R_d^{10}, the Cartesian products $\Omega_i\times R_b^{12}$ are orbits in $R_d^{10}\times R_b^{12}$. If $\hat{h}_i = \{0,d_i,0,0,1\}\in\Omega_i\times R_b^{12}$, the relations of Lemma 4.1 show that $S_{\hat{h}_i}$ is isomorphic with the stabilizer of Ω_i.

2. det$DG = 0$: We use the previous method of determination in stages of the orbits; for each orbit Ω_i of SU(2) in R_d^{10}, we choose a point $D_i^0 G\in\Omega_i$, and then we calculate in R_b^{12} the orbit X of the stabilizer of $D_i^0 G$. So, for each Ω_i, we have to discuss the conditions det $D_i^0 G = 0$ and then solve the linear system (5). In this way we finally obtain eighty-nine types of orbits, (Ω_i, X), generated by $(D_i^0 G, \times_{i=1}^{3} b_i)$, $\times_{i=1}^{3} b_i\in X$ (the interested reader may obtain their list from the authors). The corresponding stabilizers $S_{\hat{h}}$ are isomorphic to $R^{3p}\times(R^{3q}\cdot L)$ $(0\leqslant p\leqslant 4;\ 0\leqslant q\leqslant 3)$, the semidirect product being defined by the restriction of the representation $\oplus_{q=1}^{3} D_q^1$ (where $D_q^1 = D^1$, for all q) from SU(2) to L.

Q.E.D.

Proposition 4.3: The stabilizers of the semidirect products $R^{3q}\cdot L$ of the preceding proposition are isomorphic to one of the following groups: SU(2), U(1)$\cdot Z_2$, U(1), Q_8, $Q_4 Z_2$.

Remarks 4.2:

(a) All the semidirect products encountered in this method of induction in stages are regular since for each of them a Borel set can be constructed, meeting each orbit at one and only one point. So we shall obtain all the UIR of G_3 (such that the mass-operator of the Poincaré group be strictly positive) knowing those of the groups mentioned in Proposition 4.3 (which are known) by the Mackey method.

(b) If a group G is a semidirect product of two closed unimodular subgroups H and K, H being normal in G, it follows from Ref. 15 Chapter II, Paragraph 7, that G is unimodular if for every function $f\in\mathscr{L}_\mu^1(H)$ we have $\int_H f(khk^{-1})\,d\mu(h) = \int_H f(h)\,d\mu(h)$, where $d\mu$ is an invariant Haar measure on H. It follows, on applying this property in stages, that all the semidirect products encountered are unimodular and that their invariant measures are obtained by simply considering the product of the measures of their factors.

(c) The orbit $\Omega_+^{m^2}$ in R^4 obtained at the first stage of the induction and one of the orbits Ω' in $R_d^{10}\times R_b^{12}$ obtained at the second stage can be regrouped to constitute an orbit $\Omega = (\Omega_+^{m^2}, \Omega')$ in $R^4\times R_d^{10}\times R_b^{12}$ relatively to the action of $R_q^{12}\cdot$SL(2,C). This amounts to the decomposition $G_3 = H\cdot K$, where H (resp. K) is the group $R^4\times R_c^{10}\times R_a^{12}$ [resp. $R_q^{12}\cdot$SL(2,C)], and obviates the necessity for going into the details of the first two stages of the induction. Under these conditions, all the UIR of G_3 looked for are obtained (up to unitary equivalence) by choosing, for every orbit Ω of $\hat{H}$, a point $\hat{h}_0\in\Omega$, an arbitrary function of Ω in K which makes $\Gamma_{\hat{h}}$ correspond to $\hat{h}\in\Omega$ such that $\Gamma_{\hat{h}}(\hat{h}_0) = \hat{h}$ and an UIR L of the stabilizer of $\hat{h}_0$, say $K_{\hat{h}_0}$, acting in a Hilbert space $\mathscr{H}_L$. These representations are defined in the Hilbert space $\mathscr{L}_\mu^2(\Omega,\mathscr{H}_L)$ of functions F on Ω with values in $\mathscr{H}_L$ such that $\int_\Omega \|F(\hat{h})\|_{\mathscr{H}_L}^2\,d\mu(\hat{h}) < \infty$, where $d\mu(\hat{h})$ is an invariant measure under K concentrated on Ω, by

$$\{U(h_1k_1)F\}(\hat{h}) = \langle\hat{h},h_1\rangle L(\Gamma_{\hat{h}}^{-1}k_1\Gamma_{k_1\cdot\hat{h}})F(k_1^{-1}\cdot\hat{h}).$$

5. REPRESENTATIONS OF THE LIE ALGEBRA $\mathfrak{g}_3$ AND MASS SPECTRUM

In order for the interpretation that is the outcome of our model to be coherent, we must first point out the physical relevance of the generators of group G_3 and also explain the construction of the mass operators which we adopt.

Let us point out first of all that all generators of G_3 are relativistically covariant since they are invariant under translations and transform, under the action of the Lorentz group, like a Lorentz tensor. Furthermore, for each fixed pair (μ,ν), $(A_{i\mu},Q_{j\nu})$ $(1\leqslant i,j\leqslant 3)$ are "canonically conjugate" relatively to the intrinsically internal variables. So it is possible to interpret the $A_{i\mu}$ and the $Q_{j\nu}$ (and the $C_{\mu\nu}$ which depend on them algebraically) as describing the internal dynamics of a system of composite particles (the components of which may not exist in the free state) in the following way: The quark (resp. antiquark) energy–momentum will be given by an "external" part T_μ and a"internal" part $Q_{i\mu}$ (resp. $A_{i\mu}$). Similarly, we shall assume that the mass-operator for the composite system can be written (at least in a sufficient approximation) as $M^2 = M_0^2 + M_I^2$, where $M_0^2 = T_\mu T^\mu$ is the relativistic free mass and M_I^2 the mass-splitting term due to the description of particles as composite system. This mass-splitting term will be given by some combinations of the internal components $(Q_{i\mu},A_{i\mu})$ which we shall explicit in the following various cases. In what follows, M_I^2 will be called (square) mass-observable, as it is this term that, in a suitable representation of $\mathfrak{g}_3$, will give the mass-spectrum. The various mass-observables which we consider in the following are symmetric homogeneous polynomials of the second degree in the conjugate canonical variables which describe the internal dynamics. We shall thus have a description of the creation of energy from the internal motion, the particles being the excited states of an energetically more fundamental system.

In a way, this interpretation seems to be a generalization of the standard description of systems of n quantum particles possessing only properties that have a classical analog. As a matter of fact there exist, for such systems, n pairs of canonically conjugate variables $(P_i,Q_i)_{1\leqslant i\leqslant n}$ representing $\mathfrak{h}_n$ such that the Hamiltonian (and also every other physical observable) is a function of it. In our model, $\mathfrak{n}_n$, which characterizes the internal dynamics, in a first approximation, would generalize $\mathfrak{h}_n$.

This interpretation has also the advantage that none of the generators is superfluous, contrarily to the various relativistic generalizations of the SU(6) model (cf. for instance Ref. 16).

Now let us pass to the examples. We shall give four representations of the Lie algebra $\mathfrak{g}_3$. The first is integrable to the group G_3 and the three others are not, but are integrable on the Poincaré subalgebra.

A. Integrable representation of $\mathfrak{g}_3$ (harmonic oscillator model)

Let ρ be the representation of $\mathfrak{g}_3$ defined as follows: The space of the representation is the completed tensor product $\mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+,\mathscr{H}_j)\hat{\otimes}\mathscr{L}^2(R^3)$, where $\mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+,\mathscr{H}_j)$ is the carrying space of the UIR $D^+(m_0,j)$ of the Poincaré group of positive energy, mass $m_0>0$ and spin j, $\Omega^{m_0^2}_+$ is the one-sheeted hyperboloid $(p_\mu p^\mu = -m_0^2; p^4>0)$, $d\mu_0(\underline{p}) = (\vec{p}^2+m_0^2)^{-1/2}d^3\vec{p}$ is a measure concentrated on $\Omega^{m_0^2}_+$ invariant by SL(2,C), and $\mathscr{H}_j$ the carrying space of the representation of weight j of SU(2).

If we denote by the same symbol a generator (of Poincaré or of $\mathfrak{g}_3$) and its representative, the generators of $\mathfrak{g}_3$ are represented by:

$$L_{\mu\nu} = I\bar{\otimes}M_{\mu\nu},\quad T_\mu = I\bar{\otimes}P_\mu,\quad C_{\mu\nu} = I\bar{\otimes}P_\mu P_\nu,$$
$$A_{j\mu} = x_j\bar{\otimes}P_\mu,\quad Q_{j\mu} = i\left(\frac{\partial}{\partial x_j}\right)\bar{\otimes}P_\mu,$$

where I is the identity operator in $\mathscr{L}^2(R^3)$ and $\bar{\otimes}$ designates the closure of the tensor product, acting in the space $\mathscr{D}^G(D^+(m_0,j))\otimes\mathscr{S}(R^3)$, where $\mathscr{D}^G(D^+(m_0,j))$ is the Gårding domain of the representation $D^+(m_0,j)$ and $\mathscr{S}(R^3)$ is the Schwartz space.

Proposition 5.1: Let $\Omega = (\Omega^{m_0^2}_+,\Omega')$ be the orbit in $R^4\times R^{10}_d\times R^{12}_b$ characterized by the stabilized point $\hat{h}_0 = \{\underline{p}_0, D^0G, \underline{0}, \underline{0}, \underline{0}\}$ and the stabilizer $R^9\cdot$SU(2), where

$$D^0G = \begin{pmatrix} 0 & 0 \\ 0 & -m_0^2 \end{pmatrix} \text{ and } \underline{p}_0 = (0,0,0,m_0).$$

Then ρ is contained in the differential of the induced representation, starting from Ω, by the UIR of $R^9\cdot$SU(2) associated with the trivial orbit.

Proof: Any point of Ω can be parametrized by the multiple

$$\left(D(\tfrac12,\tfrac12)(A_{\underline{p}}),(\underline{p}_0),D(\tfrac12,\tfrac12)(A_{\underline{p}})D^0GD(\tfrac12,\tfrac12)(A_{\underline{p}}^{-1}),\right.$$
$$\left.\overset{3}{\underset{i=1}{\times}}D(\tfrac12,\tfrac12)(A_{\underline{p}})\begin{pmatrix}\vec{0}\\ m_0x_i\end{pmatrix}\right),$$

where $\underline{p}\to A_p$ is the field of Lorentz transformations in the canonical formalism,[13] which associates with every $\underline{p}\in\Omega^{m_0^2}_+$:

$$A_p = [2m_0(p^4+m_0)]^{-1/2}\begin{pmatrix} m_0+p^4+p^3 & p^1+ip^2 \\ p^1-ip^2 & m_0+p^4-p^3\end{pmatrix},$$
$$A_{\underline{p}}\in\mathrm{SL}(2,C) \qquad (6)$$

such that $D(\frac12,\frac12)(A_p)(\underline{p}_0) = \underline{p}$.

It follows that Ω depends on six real parameters,

$$[(p_\mu),\ \text{with}\ p_\mu p^\mu = -m_0^2,\ \text{and}\ x_i\in R\ (1\leqslant i\leqslant 3)].$$

In what follows, when clear from the context, $D(\frac12,\frac12)(A)(\underline{x})$ will simply be written $A(\underline{x})$. We can take as field $\hat{h}\to\Gamma_{\hat{h}}$ [such that $\Gamma_{\hat{h}}(\hat{h}_0) = \hat{h}$]:

$$\Gamma_{\hat{h}} = \left(\overset{3}{\underset{i=1}{\times}}A_{\underline{p}}\begin{pmatrix}\vec{0}\\ -m_0^{-1}x_i\end{pmatrix},A_{\underline{p}}\right)$$

if

$$\hat{h} = \left(\underline{p},A_pD^0GA_{\underline{p}}^{-1},\overset{3}{\underset{i=1}{\times}}A_{\underline{p}}\begin{pmatrix}\vec{0}\\ m_0x_i\end{pmatrix}\right).$$

The UIR of G_3 induced by that of $R^9\cdot$SU(2) associated with the trivial orbit in R^9 will be written

$$[U(t_0,\underline{c}_0,(\underline{a}_{0j}),(\underline{q}_{0j}),A_0F)](\underline{p},(x_i))$$
$$= \exp i[\underline{p}\cdot t_0 + \underline{d}\cdot(\underline{c}_0-\underline{\theta}_0)+\delta^{kl}b_k\cdot a_{0l}]$$
$$\times D^j(A_{\underline{p}}^{-1}A_0A_{A_0^{-1}(\underline{p})})$$
$$\times F(A_0^{-1}(\underline{p}),(x_i+m_0[A_{\underline{p}}^{-1}(\underline{q}_{0i})]^4),$$

where

$$\underline{d}^{\mu\nu} = p^\mu p^\nu,\quad b_i = x_i\underline{p},$$
$$\underline{\theta}_0^{\mu\nu} = \tfrac12\delta^{ij}[(\underline{a}_{0i}\otimes\underline{q}_{0j})^{\mu\nu}+(\underline{a}_{0i}\otimes\underline{q}_{0j})^{\nu\mu}].$$

D^j is the UIR of SU(2) of weight j operating in $\mathscr{H}_j$;

$$F\in\mathscr{L}^2_\nu(\Omega^{m_0^2}_+\times R^3;\mathscr{H}_j)\quad\text{and}\quad d\nu = d\mu_0dx_1dx_2dx_3.$$

Differentiating this representation we obtain for infinitesimal generators (defined on the space of its $\mathscr{C}^\infty$ vectors):

$$T_\mu = p_\mu,\quad A_{j\mu} = x_jp_\mu,$$
$$\vec{M} = i\vec{p}\wedge\nabla_{\vec{p}}+\vec{S},\quad Q_{j\mu} = i\left(\frac{\partial}{\partial x_j}\right)p_\mu,$$
$$\vec{N} = -(p^4+m_0)^{-1}(\vec{p}\wedge\vec{S})+ip^4\nabla_{\vec{p}},$$
$$C_{\mu\mu} = (p_\mu)^2,\quad\text{and}\quad C_{\mu\nu} = 2p_\mu p_\nu\quad(\mu\neq\nu),$$

where

$$\vec{M} = (M_1,M_2,M_3),\quad \vec{N} = (N_1,N_2,N_3),\quad M_i = \tfrac12\epsilon_{ijk}L_{jk},$$
$$N_i = L_{i4},\quad \nabla_{\vec{p}} = \left(\frac{\partial}{\partial p^1},\frac{\partial}{\partial p^2},\frac{\partial}{\partial p^3}\right),\quad \vec{S} = (S_1,S_2,S_3),$$

$(S_i)_{1\leqslant i\leqslant 3}$ are the infinitesimal generators of the representation D^j of SU(2) and $\wedge$ denotes the vector product. Q.E.D.

In this framework we shall take for the mass-splitting operator

$$M_I^2 = -g^{\mu\nu}\delta^{ij}(Q_{i\mu}Q_{j\nu}+A_{i\mu}A_{j\nu}) = M_Q^2+M_A^2. \qquad (7)$$

The internal degrees of freedom are thus assumed to contribute to the mass-splitting via a term which is (for each space–time direction) the Hamiltonian of the (three-dimensional) harmonic oscillator. This term can be viewed upon as the sum of the self-interactions of the quarks and the antiquarks, each of these self-interaction terms has a continuous spectrum, but the binding of both, described here by the sum, creates a discrete hadron mass-spectrum. This mass-observable is represented by the operator $m_0^2(-\Delta+\vec{x}^2)$ which has a discrete spectrum and leads to the mass-formula

$$m_n = m_0(2n+3)^{1/2},$$

where n is a nonnegative integer.

B. Local representations and mass-spectrum

Let us point out first of all that the use of the local (partially integrable) representations in this context is justified by the following general remark (cf. also Ref. 17): If the symmetry "differentiates" (naturally) locally, the dynamics (*a fortiori* the internal dynamics), it, on the other hand, does not necessarily "integrate" globally; in fact, direct use of the theorem of Noether leads (by differentiation) to observables obeying laws of conservation; conversely, if we start out from an algebra of observables, each of them can generate a one-parameter group, but the dynamics by no means requires that there is global integrability; now in a situation such as this, it is the dynamics which must take precedence. As for the method of investigation of the local representations of $\mathfrak{g}_3$, it is analogous to that Ref. 18, the definitions of which we adopt. They are constructed upon the differential of a unitary global representation of $\mathfrak{g}_3$, carried out in a certain functional space, by truncating the domain of variation of the variables with suitable boundary conditions in order to destroy integrability. These boundary conditions can also be considered as reflecting the internal dynamics (or even the internal geometry) of hadrons.

We are going to study here three irreducible Poincaré partially integrable local representations ρ^i $(1\leqslant i\leqslant 3)$ of the Lie algebra $\mathfrak{g}_3$, related to the integrable representation ρ defined above. By irreducibility of ρ^i we understand Schur-irreducibility, namely that every bounded operator commuting strongly with some integrable observables (i.e., with their spectral resolutions) and weakly with the others is a multiple of the identity operator. More precisely we require the commutation with the unitary group corresponding to maximum integrable Lie subalgebra of $\rho^i[\mathcal{U}(\mathfrak{g}_3)]$ and commutation with the operators of $\rho^i[\mathcal{U}(\mathfrak{g}_3)]$ on the common invariant domain. Note that the integrable Lie subalgebra of $\rho^i[\mathcal{U}(\mathfrak{g}_3)]$ need not be the representation of a Lie subalgebra of $\mathcal{U}(\mathfrak{g}_3)$, since (as will happen for ρ^2 and ρ^3) some elements of $\mathcal{U}(\mathfrak{g}_3)$ can be trivially represented (i.e., by multiples of the identity). In the examples that we consider it will be enough to restrict ourselves to elements of degree less than or equal to two in $\rho^i[\mathcal{U}(\mathfrak{g}_3)]$. For each ρ^i, we introduce two domains: The domain of definition of the representation, and the mass-spectrum domain on which the mass-observable considered is represented by an essentially self-adjoint positive operator with a purely discrete spectrum, consisting of isolated eigenvalues. The former is a subspace of the latter.

(a) Let $\mathscr{L}^1=\mathscr{L}^2_{\mu 0}(\Omega^{m_0^2}_+,\mathscr{H}_j)\,\hat{\otimes}\,\mathscr{L}^2(Q)$, where Q is the cube $(0\leqslant x_i\leqslant a)$. We write $S^1_0=\mathscr{S}_{\vec{p}}(R^3,\mathscr{H}_j)\otimes\mathscr{C}^\infty_0(Q)$, where $\mathscr{S}_{\vec{p}}(R^3,\mathscr{H}_j)$ is the space of the rapidly decreasing $\mathscr{C}^\infty$ (in $\vec{p}$) functions of $\mathscr{L}^2_{\mu 0}(\Omega^{m_0^2}_+,\mathscr{H}_j)$ and $\mathscr{C}^\infty_0(Q)$ is the space of $\mathscr{C}^\infty$ (in $\vec{x}$) functions, vanishing together with all their derivatives on the boundary of Q. S^1_0 is dense in $\mathscr{H}^1$ and is a common invariant domain of the infinitesimal generators of the representation ρ on which they are symmetric. This defines a representation ρ^1 of $\mathfrak{g}_3$ on $\mathscr{H}^1$ by symmetric operators.

Now, let B be an arbitrary bounded operator. We shall make use of the strong commutation with the elements $\rho^1(L_{\mu\nu})$, $\rho^1(T_\mu)$, $\rho^1(g^{\mu\nu}T_\mu Q_{j\nu})$ which define on $S^1_\pi=\mathscr{S}_{\vec{p}}(R^3,\mathscr{H}_j)\otimes\mathscr{C}^\infty_\pi(Q)$, where $\mathscr{C}^\infty_\pi(Q)$ is the space of the $\mathscr{C}^\infty$ (in $\vec{x}$) functions periodic on the boundary of Q together with all their derivatives, an integrable representation[17] of this subalgebra $R^3\oplus[R^4\oplus\mathfrak{sl}(2,C)]$. This implies that B must satisfy

$$B\left\{\psi_l(\vec{p},s)\exp\left(i\frac{2\pi}{a}\vec{n}\vec{x}\right)\right\}=b(\vec{n})\psi_l(\vec{p},s)\exp\left(i\frac{2\pi}{a}\vec{n}\vec{x}\right),$$

where $[\psi_l(\vec{p},s)$ denoting a basis of $\mathscr{L}^2_{\mu 0}(\Omega^{m_0^2}_+,\mathscr{H}_j)]$

$$\left\{\psi_l(\vec{p},s)\exp\left(i\frac{2\pi}{a}\vec{n}\vec{x}\right),l\in N,\ -j\leqslant s\leqslant j,\vec{n}\in Z^3\right\}\subset S^1_\pi$$

is a basis of $\mathscr{H}^1$. Finally the weak commutation with $\rho^1(g^{\mu\nu}T_\mu A_{j\nu})$ and the density of S^1_0 in $\mathscr{H}^1$ show that B must be scalar. So ρ^1 is a Schur-irreducible representation of $\mathfrak{g}_3$.

Denoting by V the unitary transformation on $\mathscr{H}^1$ which maps f on $\exp(\frac{1}{2}i\vec{x}^2)f$, $V(S^1_\pi)$ contains an invariant dense common domain of analytic vectors for each of the generators $L_{\mu\nu}$, T_μ, $\Sigma^3_{j=1}(Q_{j\mu}+A_{j\mu})$ which is:

$$\mathscr{D}^\omega=\{\psi e^{i\vec{x}^2/2}e^{i2\pi\vec{n}\cdot\vec{x}/a};\psi\in\mathscr{D}^\omega(D^+(m_0j)),\vec{n}\in Z^3\}$$

where $\mathscr{D}^\omega(D^+(m_0j))\subset\mathscr{S}_{\vec{p}}(R^3,\mathscr{H}_j)$ is the set of analytic vectors corresponding to the representation $D^+(m_0j)$ of $\mathfrak{p}$. Then (see Ref. 17) these generators define on $V(S^1_\pi)$ a representation of $R^8\oplus\mathfrak{sl}(2,C)$ integrable to a unitary representation of the group $R^8\cdot\mathrm{SL}(2,C)$.

The squared mass-operator

$$M^2=-g^{\mu\nu}\left\{\sum_{j=1}^3(Q_{j\mu}+A_{j\mu})\right\}\left\{\sum_{j=1}^3(Q_{j\nu}+A_{j\nu})\right\}\qquad(8)$$

is an invariant of $R^8\cdot\mathrm{SL}(2,C)$. It is represented in this representation ρ^1 by $m_0^2\{\Sigma^3_{j=1}(i\partial_j+x_j)\}^2$, which is essentially self-adjoint on $V(S^1_\pi)$ and has a discrete spectrum which leads to the mass formula

$$m_k=2\pi ka^{-1}m_0,$$

where k is a positive integer.

(b) Let $\mathscr{H}^2=\mathscr{L}^2_{\mu 0}(\Omega^{m_0^2}_+,\mathscr{H}_j)\,\hat{\otimes}\,\mathscr{L}^2(C)$, where C is the cylinder

$$\{(x_1,x_2,x_3);0\leqslant x_1^2+x_2^2\leqslant a^2\ \text{ and }\ 0\leqslant x_3\leqslant b\}.$$

$\mathscr{L}^2(C)$ decomposes into

$$\mathscr{L}^2(C)=\mathscr{L}^2([0,a]),rdr)\,\hat{\otimes}\,\mathscr{L}^2(T,d\theta)\,\hat{\otimes}\,\mathscr{L}^2([0,b],dx_3),$$

where T is the torus $[0,2\pi]$ and $x_1=r\cos\theta$, $x_2=r\sin\theta$. Let $S^2_0=\mathscr{S}_{\vec{p}}(R^3,\mathscr{H}_j)\otimes\mathscr{C}^\infty_{0r}([0,a])\otimes\mathscr{C}^\infty_\theta(T)\otimes\mathscr{C}^\infty_{0x_3}([0,b])$ be, where $\mathscr{C}^\infty_{0r}([0,a])$ and $\mathscr{C}^\infty_{0x_3}([0,b])$ are the spaces of the functions $\mathscr{C}^\infty$ (in r or x_3 respectively) vanishing together with all derivatives at the end points of the intervals, and $\mathscr{C}^\infty_\theta(T)$ that of the $\mathscr{C}^\infty$-functions on $T\cdot S^2_0$, which is dense in $\mathscr{H}^2$, is a common invariant domain of the infinitesimal generators of the representation ρ, on which they are symmetric. This defines a local representation ρ^2 of $\mathfrak{g}_3$ on $\mathscr{H}^2$ by symmetric operators.

Now, let B be an arbitrary bounded operator which commutes strongly with $\rho^2(L_{\mu\nu})$, $\rho^2(T_\mu)$, and $\rho^2(g^{\mu\nu}T_\mu Q_{3\nu})$

which generate the group $P \times U(1)$ and weakly with ρ^2 $(g^{\mu\nu}T_\mu A_{3\nu})$. Then B reduces to

$$B\psi_l(\vec{p},s)J_n(k_n r)e^{\pm in\theta}e^{i2\pi z x_3/b} = \psi_l(\vec{p},s)e^{i2\pi z x_3/b}B' J_n(k_n r)e^{\pm in\theta},$$

where

$$\{\psi_l(\vec{p},s)J_n(k_n r)e^{\pm in\theta}e^{i2\pi z x_3/b} \text{ with } l\in N, -j\leqslant s\leqslant j, n\in N, z\in Z, k_n \text{ verifying } J_n(k_n a)=0\}$$

is a basis of $\mathscr{H}^2$, J_n being the Bessel functions, solution to the Sturm–Liouville problem relative to the vibrating circular membrane.[19] The two operators $\rho^2(\Sigma_{j=1}^2 g^{\mu\nu}Q_{j\mu}Q_{j\nu})$ and $\rho^2(g^{\mu\nu}(Q_{1\mu}A_{2\nu} - Q_{2\mu}A_{1\nu}))$ generate unitary representations of U(1) commuting between themselves and with the group $P \times U(1)$ represented above. Our assumption of Schur-irreducibility implies that B must commute with these groups also, and therefore

$$B' J_n(k_n r)e^{\pm in\theta} = b(\pm n, k_n)J_n(k_n r)e^{\pm in\theta}.$$

Finally an argument analogous to that of (Ref. 20 Appendix) using the weak commutation with $\rho^2(g^{\mu\nu}T_\mu Q_{1\nu})$ shows that B is scalar. So ρ^2 is a Schur-irreducible local representation of $\mathfrak{g}_3$ on $\mathscr{H}^2$.

Assuming (at least in a first approximation) that the "internal momenta" $Q_{i\mu}$ of the "quarks" and $A_{i\mu}$ of the "antiquarks" add linearly and that there is no contribution mixing the different "colors" $i = 1,2,3$, we are led to consider the mass-splitting operator,

$$M_I^2 = -g^{\mu\nu}\delta^{ij}(Q_{i\mu} + A_{i\mu})(Q_{j\nu} + A_{j\nu}). \tag{9}$$

It is represented, in the representation ρ^2, by the operator

$$M_I^2 = m_0^2 \sum_{j=1}^{3} (i\partial_j + x_j)^2 \tag{9a}$$

which is essentially self-adjoint on the domain $V(S_\pi^2)$, where

$$S_\pi^2 = \mathscr{L}_\mu^2(\Omega_+^{m_0^2}, \mathscr{H}_j) \otimes \mathscr{C}_{0r}^{'\infty}([0,a]) \otimes \mathscr{C}_\theta^\infty(T) \otimes \mathscr{C}_{\pi x_3}^\infty([0,b]),$$

$\mathscr{C}_{0r}^{'\infty}([0,a])$ being the space of the functions $\mathscr{C}^\infty$ (in r), vanishing at the right endpoint a, and $\mathscr{C}_{\pi x_3}^\infty([0,b])$ being that of the functions $\mathscr{C}^\infty$ in x_3 and verifying, together with all their derivatives, the boundary condition of periodicity. M^2 has a discrete spectrum which leads to the mass formula

$$m_{n,k_n,z} = m_0[k_n^2 + (2\pi z b^{-1})^2]^{1/2},$$

where $n\in N$, $k_n\in S_n = \{\lambda; J_n(\lambda a) = 0\}$, and $z\in Z$ (all integers). This model could be generalized by introducing, instead of δ^{ij}, different tensors τ^{ij} in (9) which by their nature would give rise to internal quantum numbers, for instance.

(c) Let $\mathscr{H}^3 = \mathscr{L}_{\mu_0}^2(\Omega_+^{m_0^2}, \mathscr{H}_j)\,\widehat{\otimes}\,\mathscr{L}^2(S)$, where S is the sphere $\{\vec{x}\in R^3 \text{ such that } |\vec{x}|\leqslant a\}$. $\mathscr{L}^2(S)$ decomposes into

$$\mathscr{L}^2(S) = \mathscr{L}^2([0,a], r^2dr)\,\widehat{\otimes}\sum_{j\in N}\mathscr{H}_j(\theta,\varphi),$$

where (r,θ,φ) are the spherical coordinates and $\mathscr{H}_j(\theta,\varphi)$ is the carrying space of the UIR of weight j of SO(3). Let S_0^3 be the dense subspace of $\mathscr{H}^3$ defined by

$$S_0^3 = \mathscr{S}_\beta(R^3, \mathscr{H}_j) \otimes \mathscr{C}_{0r}^\infty([0,a]) \otimes \sum_{j\in N}\mathscr{H}_j(\theta,\varphi).$$

S_0^3 is a common invariant domain of the infinitesimal generators of the representation ρ on which they are symmetric.

This defines a local representation ρ^3 of $\mathfrak{g}_3$ on $\mathscr{H}^3$ by symmetric operators. The proof of the Schur-irreducibility is analogous to that of the preceding case. First we consider a basis of $\mathscr{H}^3$, defined by the following functions:

$$\psi_n(\vec{p},s)r^{-1/2}J_{l+1/2}(k_l r)Y_{m_l}^l(\theta,\varphi),$$

$l\in N$, $-l\leqslant m_l\leqslant l$, and $k_l\in S_l = \{\lambda; J_{l+1/2}(\lambda a) = 0\}$.

The $J_{l+1/2}$ are the Bessel functions, solution to the Sturm–Liouville problem relative to the vibrating sphere,[19] and the $Y_{m_l}^l(\theta,\varphi)$ are the spherical harmonics. Now we use the strong commutation with $\rho^3(L_{\mu\nu})$, $\rho^3(T_\mu)$, $\rho^3(g^{\mu\nu}\delta^{ij}Q_{i\mu}Q_{j\nu})$, which generates a $P \times U(1)$ group and $\rho^3(L_{ij})$, where $L_{ij} = g^{\mu\nu}(Q_{i\mu}A_{j\nu} - Q_{j\mu}A_{i\nu})$ $(1\leqslant i,j\leqslant 3)$. The L_{ij} do not close to a $\mathfrak{so}(3)$ subalgebra of $\mathscr{U}(\mathfrak{g}_3)$; but since $\rho^3(g^{\mu\nu}T_\mu T_\nu)$ factors out in $\rho^3(L_{ij})$ and is a multiple of identity, the $\rho^3(L_{ij})$ generate the quasiregular representation of SO(3) in the $\vec{x}$-part of $\mathscr{H}^3$. We therefore require that B commutes with the above unitary representation of $P \times U(1) \times SO(3)$; weak commutation with $\rho^3(g^{\mu\nu}T_\mu Q_{3\nu})$ will then imply that B is a multiple of identity. Let

$$S_\pi^3 = \mathscr{L}_\mu^2(\Omega_+^{m_0^2}, \mathscr{H}_j) \otimes \mathscr{C}_{0r}^{'\infty}([0,a]) \otimes \sum_{l\in N}\mathscr{H}_l(\theta,\varphi).$$

The mass-splitting operator M_I^2 defined by (9a) is essentially self-adjoint on the domain $V(S_\pi^3)$, where V is the previously defined unitary operator. It has discrete spectrum which leads to the mass-formula

$$m_{l,k_l} = m_0|k_l| \quad (l\in N, k_l\in S_l).$$

Remark: Using also a three-dimensional "internal space" but a local representation of the 11-parameter Weyl Lie algebra, and taking for total (squared) mass-operator the second order invariant of the (integrable) Poincaré subalgebra, Snellman[20] recently obtained a discrete mass-spectrum in terms of squares of the zeros of Bessel functions relative to the vibrating sphere. The technique used for ρ^2 and ρ^3 is similar, but these representations are Schur-irreducible and our interpretation is different.

6. POSSIBLE INTRODUCTION OF THE ISOSPIN

Let G_3-$SU_I(2)$ be the Lie group whose Lie algebra is $U(\mathfrak{g}_3, \mathfrak{su}_I(2))$ defined by the method indicated in Remark 2.1(b). $SU_I(2)$, commuting with the external symmetry, will be interpreted as being the isospin group. As for the relativistic spin, we shall continue in this article to take as our definition the one introduced by Bargmann and Wigner (cf., for example, Ref. 13). In this interpretation, the spin remains linked integrally with the external symmetry; but the mass, while depending on the external symmetry, is linked with the degrees of internal excitation of our composite model. In this section we are going to determine the isopin content of the two interaction operators introduced in Sec. 5 A and 5 B (c), the associated models of which shall, in what follows, be called respectively the harmonic oscillator model and the

vibrating sphere model. A glance at these operators M^2 and formula (10) below easily shows that the eigenfunctions of M^2 will be classified by a principal quantum number connected with the internal excitation level and a secondary quantum number associated with the weight of an UIR of $SU_I(2)$.

A. Harmonic oscillator model

Let $\mathscr{H} = \mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+, \mathscr{H}_j) \hat{\otimes} \mathscr{L}^2(R^3) \hat{\otimes} \mathscr{H}_k$, where $\mathscr{H}_k$ is the carrying space of the representation of weight k of $SU_I(2)$. In $\mathscr{H}$, we can define a representation ω of $\mathfrak{g}_3 \oplus \mathfrak{su}_I(2)$ by symmetric operators, which is Schur-irreducible and integrable to $G_3 \cdot SU_I(2)$, as follows: The common invariant dense domain of the infinitesimal generators of ω is

$$\mathscr{D}^G(D^+(m_0 j)) \otimes \mathscr{S}(R^3) \otimes \mathscr{H}_k$$

on which the generators of $\mathfrak{g}_3$ are represented by the basis of the representation ρ of Proposition 5.1 and those of $\mathfrak{su}_I(2)$, namely $\vec{I} = (I_1, I_2, I_3) = (I_{23}, I_{31}, I_{12})$, are represented by

$$\vec{I} = \vec{\sigma}^k + i\vec{x} \wedge \nabla_{\vec{x}}, \tag{10}$$

where $\vec{\sigma}^k = (\sigma_1^k, \sigma_2^k, \sigma_3^k)$ is the basis of the representation D^k of $\mathfrak{su}_I$ and $\nabla_{\vec{x}} = (\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$.

Let $\mathscr{H}_n$ be the eigensubspace associated with the eigenvalue $m^2 = m_0^2(2n+3)$ of the interaction operator $-g^{\mu\nu}\delta^{ij}(Q_{i\mu}Q_{j\nu} + A_{i\mu}A_{j\nu})$ which is represented in ω by $M^2 = m_0^2(-\Delta + \vec{x}^2)$. $\mathscr{H}_n$ is written

$$\mathscr{H}_n = \mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+, \mathscr{H}_j) \hat{\otimes} E(n) \hat{\otimes} \mathscr{H}_k,$$

where $E(n)$ is the eigensubspace associated with the eigenvalue $\lambda_n = (2n+3)$ of the operator $-\Delta + \vec{x}^2$.

As $SU_I(2)$ does not operate in $\mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+, \mathscr{H}_j)$, all we need to know in order to elucidate the isospin content of $\mathscr{H}_n$ is the decomposition of the unitary representation ω_I of $SU_I(2)$ in $E(n) \hat{\otimes} \mathscr{H}_k$ into irreducible components. Now λ_n represents the energy-spectrum of the three-dimensional harmonic oscillator, so the degeneracy of the level λ_n is equal to the dimension of the representation $(n,0)$ in the notation of Ref. 21) of the group SU(3), which implies that $\dim E(n) = (n+1)(n+2)/2$. Starting from the decomposition of the restriction of the representation $(n,0)$ from SU(3) to SO(3),[22] we obtain:

$$\omega_I = \sum_{s=0}^{n/2} \sum_{i=|n-2s-k|}^{n-2s+k} D^i, \quad \text{if } n \text{ is even},$$

$$\omega_I = \sum_{s=0}^{(n-1)/2} \sum_{i=|n-2s-k|}^{n-2s+k} D^i, \quad \text{if } n \text{ is odd}.$$

As for the spin states of the eigenvectors of M^2, they are all equal to j, as the restriction of the representation ω to the Poincaré group decomposes on $\mathscr{H}_n$ according to $\sum_{\alpha=1}^{(n+1)(n+2)(2k+1)/2} D^+_\alpha(m_0 j)$, where $D^+_\alpha(m_0 j) = D^+(m_0 j)$ for all α.

Let us detail the cases $k = 0, \frac{1}{2}$. To a mass $m^2 = m_0^2(2n+3)$ correspond:

(a) if $k = 0$, $(n+2)/2$ [resp. $(n+1)/2$] particles with isospin $n-2s$, where $s = 0,1,\ldots,n/2$ [resp. $s = 0,1,\ldots,(n-1)/2$] if n is even (resp. odd), each one having the same spin j;

(b) if $k = \frac{1}{2}$, $(n+1)$ particles with isospin $(2s+1)/2$, where $s = 0,1,\ldots,n$ and the same spin j.

B. Vibrating sphere model

Let $\mathscr{H} = \mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+, \mathscr{H}_j) \hat{\otimes} \mathscr{L}^2(S) \hat{\otimes} \mathscr{H}_k$. In $\mathscr{H}$, we can define a representation ω of $\mathfrak{g}_3 \oplus \mathfrak{su}_I(2)$, by symmetric operators and Schur-irreducible, as follows: The common invariant dense domain of the infinitesimal generators of ω is $S_0^3 \otimes \mathscr{H}_k$ on which the generators of $\mathfrak{g}_3 \oplus \mathfrak{su}_I(2)$ are represented as in Sec. 6 A.

Let $\mathscr{H}(s,k_s)$ be the eigensubspace associated with the eigenvalue $m^2 = m_0^2 k_s^2$ of the interaction operator $-g^{\mu\nu}\delta^{ij}(Q_{i\mu} + A_{i\mu})(Q_{j\nu} + A_{j\nu})$ which is represented in ω by $M^2 = m_0^2 \Sigma_{j=1}^3 (i\partial_j + x_j)^2$. $\mathscr{H}(s,k_s)$ is written

$$\mathscr{H}(s,k_s) = \mathscr{L}^2_{\mu_0}(\Omega^{m_0^2}_+, \mathscr{H}_j) \hat{\otimes} E(s,k_s) \hat{\otimes} \mathscr{H}_k,$$

where $E(s,k_s)$ is the subspace of $\mathscr{L}^2(S)$ generated by the vectors

$$r^{-1/2} \exp i(r^2/2) J_{s+1/2}(k_s r) Y^s_{m_s}(\theta,\varphi), \quad -s \leqslant m_s \leqslant s.$$

As in the preceding case, in order to elucidate the isospin content of $\mathscr{H}(s,k_s)$ it is enough to know the decomposition of the unitary representation ω_I of SU(2) in $E(s,k_s) \hat{\otimes} \mathscr{H}_k$. Now $\omega_I = \Sigma_{i=|s-k|}^{s+k} D^i$. So the states of $\mathscr{H}(s,k_s)$ are all of spin j and mass $m_0|k_s|$ (k_s being determinable approximatively for large masses using an asymptotic representation of the Bessel functions) and their isospin (for the particular values $k = 0, \frac{1}{2}$, for example) is:

(a) $I = s$, if $k = 0$,

(b) $I = \frac{1}{2}$ (resp. $I = s \pm \frac{1}{2}$ if $s = 0$ (resp. $s \geqslant 1$) and $k = \frac{1}{2}$.

7. DISCUSSION AND OUTLOOK

After having traced the experimental trajectories $I = f(m^2)$ and $I = f(m)$ (with constant spin) for those particles which are best established at present[23] it emerges that our model presents a real potential interest, as the observed divergences from the theoretical trajectories (both those associated with the harmonic oscillator model and those associated with the vibrating sphere model) are not significant enough to invalidate it; all the more so, as we have only taken into account a restricted number of quantum numbers. The presence of high isospins which do not correspond to observed particles could be accounted for by the fact that they are associated with particles heavy enough to decay very quickly by means of strong interactions. Furthermore, the fact that this model permits defining a denumerable infinity of fermions (or of bosons) of integer or half-integer isospin (in an irreducible representation) brings up the question of the interest of the "multiquark" theories ("colored," "charmed,"···), especially as there are some very strong presumptions as to the existence of massive hadrons which do not fit into the framework of the present multiquark theories.

We end this work by mentioning briefly a few remarks and suggestions concerning its possible continuations:

(1) The relation $\vec{I} = \vec{\sigma}^{\,k} + i\vec{x} \wedge \nabla_{\vec{x}}$ giving the isospin can be interpreted as follows: $\vec{\sigma}^{k}$ represents the isospin of the basic (nonexcited) state and $i\vec{x} \wedge \nabla_{\vec{x}}$ is the orbital momentum of internal excitation creating the real isopin. If a mass formula were desired depending explicitely on the isospin one could add to M_I^2 an interaction of the (iso) spin-orbit type of the form $I_{ij}\Sigma^{ij}$, where Σ^{ij} is an appropriate function of the internal canonical variables $(A_{i\mu},Q_{j\nu})$. For instance, if we take the mass-observable $M^2 = M_I^2 - gI_{ij}\Sigma^{ij}$ (g being a coupling constant) with $\Sigma_{ij} = g^{\mu\nu}(A_{i\mu}Q_{j\nu} - A_{j\mu}Q_{i\nu})$, we obtain, in the harmonic oscillator framework, the following mass-formula

$$m^2 = m_0^2(2n+3) + gm_0^2[I(I+1) + l(l+1) - k(k+1)],$$

with

$$l = n - 2s, \quad 0 \leqslant s \leqslant \text{integer part of } n/2,$$
$$I \in \{n - 2s + k, n - 2s + k - 1, \ldots, |n - 2s - k|\}.$$

(2) In another perspective it would be interesting to define (working from internal canonical variables) operators $X_{i\mu}^{\pm}$ of the creation and annihilation type, such that the internal quantum numbers (including the isospin) be functions of them, and to deduce from the latter the internal quantum number content of the various mass-formulas considered (as we have done above for the isospin), or mass-formulas which depend on them explicitly as in the preceding paragraph.

(3) Certain subalgebras of $\mathfrak{g}_3$ (or of $\mathfrak{g}_3 \oplus \mathfrak{l}_3$) the deformations of which we have studied (such as, for instance, $R^5 \oplus \mathfrak{su}(2)$ and $R^{12} \oplus (\mathfrak{sl}(2,C) \oplus \mathfrak{l}_3)$) are likely to contain a description of the internal quantum numbers. The study of their behavior in the representations considered could lead to an elucidation, from another viewpoint, of the internal quantum number content of the various mass-formulas proposed.

(4) It would be interesting to study other local representations of $\mathfrak{g}_3$, in particular those deducible from the differentials of the induced global representations, the inducing subgroup of which (cf. Proposition 4.2) admits SU(2) as homogeneous factor (which is linked with the spin), such as the ones reported in the Appendix. It is probable that for some of them, contrary to those studied in this paper, a spin spectrum might be found which would lead to a group theoretical formulation of the Regge trajectories. More simply, in view of the results obtained in this paper, the fact that the group $H_4 \cdot SL(2,C)$[24] admits (nonfaithful) UIR such that the Poincaré mass is a strictly positive scalar and such that the spin-spectrum is composed of all the integers or all the half-integers presents another possibility for arriving at a formulation of Regge trajectories by building representations of $\mathfrak{g}_3$ [or $\mathfrak{g}_3 \oplus \mathfrak{su}_I(2)$] according to our analysis (cf. Sec. 2) for which the restriction to the Poincaré subalgebra gives a spin-spectrum, as in the Flato–Snellman model (Ref. 25, Sec. 2 C).

(5) Let us mention in conclusion that the methods employed in this paper could equally be considered in the framework of the algebras of supersymmetries by adjoining anticommutation relations to the commutation relations of $\mathfrak{h}_n$.

ACKNOWLEDGMENTS

Our study of this problem was motivated by ideas of M. Flato. We wish to thank him for his friendly help and criticism on the topics of this work. We should like also to thank D. Sternheimer for his careful reading of the manuscript and useful remarks.

APPENDIX

We construct here the differentials of the UIR of G_3 associated with the two classes of orbits whose stabilizers are SU(2) and $R^3 \times SU(2)$ respectively. These representations might be useful for a spin-spectrum. The notations are those of Sec. 5 A.

(a) Let $\Omega = (\Omega_+^{m^2}, \Omega(\lambda,\lambda,\lambda) \times \{0\} \times \{\alpha\} \times R_b^{12})$ be the orbit whose stabilizer is SU(2) and whose stabilized point is $\hat{h}_0 = (\underline{p}_0, D^0G, \underline{0}, \underline{0}, 0)$. Any point $\hat{h}$ of Ω can be parametrized by the multiple

$$(A_{\underline{p}}(\underline{p}_0), A_{\underline{p}} D^0 G A_{\underline{p}}^{-1}, \overset{3}{\underset{i=1}{\times}} b_i),$$

where $\underline{p}_0 = (0,0,0,m)$, $\underline{p} \to A_{\underline{p}}$ is the field of Lorentz transformations (6)

$$D^0G = \begin{pmatrix} [\lambda] & 0 \\ 0 & -\alpha \end{pmatrix},$$

where $[\lambda]$ is the scalar matrix defined by λ. So Ω depends on the following fifteen parameters:

(p^μ with $p_\mu p^\mu = -m^2, b_i^\mu \in R$; $1 \leqslant i \leqslant 3, 1 \leqslant \mu \leqslant 4$).

The field $\hat{h} \to \Gamma_{\hat{h}}$ of Ω in $R_q^{12} \cdot SL(2,C)$, which permits passing from the stabilized point $\hat{h}_0$ to any point $\hat{h}$ of Ω, can be written

$$\Gamma_{\hat{h}} = \left(\overset{3}{\underset{i=1}{\times}} A_{\underline{p}} D^0 G^{-1} A_{\underline{p}}^{-1} b_i, A_{\underline{p}} \right).$$

The representation of G_3 induced by the representation D^j (operating in $\mathscr{H}_j$) of the inducing subgroup SU(2) can be written:

$$\{U(t_0,c_0,(a_{0i}),(q_{0i}),A_0)F\}(\underline{p},(b_i)_{1 \leqslant i \leqslant 3})$$
$$= \exp i\{\underline{p}\cdot t_0 + \underline{d}\cdot(\underline{c}_0 - \underline{\theta}_0) + \delta^{k1} b_k \cdot a_{01}\}$$
$$\times D^j(A_{\underline{p}}^{-1} A_0 A_{\Lambda_{A_0}^{-1}(p)}) F(A_0^{-1}\underline{p},$$
$$\times (A_0^{-1}(b_i - DGq_{0i}))_{1 \leqslant i \leqslant 3}),$$

where $DG = A_{\underline{p}} D^0 G A_{\underline{p}}^{-1}$, $\underline{d}$ is the vector of R_d^{10} defined by DG. The space of the representation is $\mathscr{L}_\nu^2(\Omega_+^{m^2} \times R_b^{12}, \mathscr{H}_j)$ where

$$d\nu = (\vec{p}^2 + m^2)^{-1/2} d^3\vec{p} \prod_{\substack{1 \leqslant i \leqslant 3 \\ 1 \leqslant \mu \leqslant 4}} db_i^\mu.$$

A basis of the differential (operating on the space of the $\mathscr{C}^\infty$-vectors) of the representation U is given by:

$$T_\mu = p_\mu, \quad A_{j\mu} = b_{j\mu}, \quad C_{\mu\nu} = d_{\mu\nu},$$
$$(Q_{j1}, Q_{j2} Q_{j3}, Q_{j4}) = i\left(\frac{\partial}{\partial b_j^1}, \frac{\partial}{\partial b_j^2}, \frac{\partial}{\partial b_j^3}; \frac{\partial}{\partial b_j^4}\right) DG,$$
$$\vec{M} = \vec{S} + i\{(\vec{p} \wedge \nabla_{\vec{p}}) + \delta^{kl}(\vec{b}_k \wedge \nabla_{\vec{b}_l})\},$$

$$\vec{N} = -(p^4+m)^{-1}(\vec{p}\wedge\vec{S}) + i\{p^4\nabla_{\vec{p}} + \vec{B}\},$$

where $\vec{B}$ is the 3-vector of components

$$B^j = \delta^{kl}\left(b^4_k\frac{\partial}{\partial b^j_l} + b^j_k\frac{\partial}{\partial b^4_l}\right), \quad 1\leqslant j\leqslant 3.$$

(b) Let $\Omega = (\Omega^{m^2}_{+}, \Omega')$ be the orbit whose stabilizer is

$$R^3\times \mathrm{SU}(2) = \left\{\left(\overset{3}{\underset{i=1}{\times}}\begin{pmatrix}\vec{0}\\ q^4_1\end{pmatrix}, A\right);\ A\in\mathrm{SU}(2)\right\}$$

and whose stabilized point is

$$\hat{h}_0 = \left(\underline{p}_0, D^0G, \overset{3}{\underset{i=1}{\times}}\begin{pmatrix}\vec{0}\\ \beta_i\end{pmatrix}\right), \text{ where } D^0G = \begin{pmatrix}[\lambda\ ] & 0\\ 0 & 0\end{pmatrix}.$$

Any point $\hat{h}$ of Ω can be parametrized by the multiple

$$\left(A_{\underline{p}}(\underline{p}_0), A_{\underline{p}}D^0GA^{-1}_{\underline{p}}, \overset{3}{\underset{i=1}{\times}}A_{\underline{p}}\begin{pmatrix}\lambda\vec{y}_i\\ \beta_i\end{pmatrix}\right),$$

where $\vec{y}_i\in R^3$ $(1\leqslant i\leqslant 3)$. So Ω depends on the twelve parameters (p^μ such that $p_\mu p^\mu = -m^2$, $y^j_i\in R$, $1\leqslant i, j\leqslant 3$). Let us take the field of transformations

$$\hat{h}\to\Gamma_{\hat{h}} = \left(\overset{3}{\underset{i=1}{\times}}A_{\underline{p}}\begin{pmatrix}\vec{y}_i\\ 0\end{pmatrix},\ A_{\underline{p}}\right).$$

The representation of G_3, induced by the UIR of $R_3\times\mathrm{SU}(2)$: $(\vec{x},A)\to\exp i(\vec{r}\cdot\vec{x})D^j(A)$, where $\vec{x},\vec{r}\in R^3$ and $A\in\mathrm{SU}(2)$, can be written

$$\{U\{t_0,\underline{c}_0,(\underline{a}_{0i}),(\underline{q}_{0i}),A_0\}F\}(\underline{p};(\vec{y}_i)_{1\leqslant i\leqslant 3})$$
$$= \exp i\left\{\underline{p}\cdot t_0 + \underline{d}\cdot(\underline{c}_0 - \underline{\theta}_0) + \delta^{k\,1}\underline{b}_k\cdot\underline{a}_{01} + \sum_{k=1}^{3} r_k[A^{-1}_{\underline{p}}\underline{q}_{0k}]^4\right\}$$
$$\times D^j(A^{-1}_{\underline{p}}A_0A_{A_{\cdot\cdot}'(\underline{p})})F(A_0^{-1}\underline{p},(\vec{y}'_i)_{1\leqslant i\leqslant 3})$$

where

$$b_k = A_{\underline{p}}\begin{pmatrix}\lambda\vec{y}_k\\ \beta_k\end{pmatrix}, \quad \vec{y}'_i = D^1(A^{-1}_{\underline{p}}A_0A_{A_{\cdot\cdot}'(\underline{p})})^{-1}(\vec{y}_i - \overrightarrow{A^{-1}_{\underline{p}}\underline{q}_{0i}}).$$

The space of the representation is $\mathscr{L}^2_\nu(\Omega^{m^2}_{+}\times R^9, \mathscr{H}_j)$, where

$$d\nu = (\vec{p}^2+m^2)^{-1/2}d^3\vec{p}\prod_{i=1}^{3}d^3\vec{y}_i.$$

A basis of the differential (operating on the space of the $\mathscr{C}^\infty$-vectors) of this representation U is given by:

$$T_\mu = p_\mu, \quad A_{j\mu} = b_{j\mu}, \quad C_{\mu\nu} = d_{\mu\nu},$$
$$Q_{jk} = -m^{-1}r_jp_k + i\left(\frac{\partial}{\partial y^k_j}\right) + im^{-1}(p^4+m)^{-1}p^k\sum_{l=1}^{3}p^l\left(\frac{\partial}{\partial y^l_j}\right), \quad 1\leqslant k\leqslant 3,$$
$$Q_{j4} = -m^{-1}\left\{r_jp^4 + i\sum_{l=1}^{3}p^l\left(\frac{\partial}{\partial y^l_j}\right)\right\},$$
$$\vec{M} = \vec{S} + i\{\vec{p}\wedge\nabla_{\vec{p}} + \delta^{kl}(\vec{y}_k\wedge\nabla_{\vec{y}_l})\},$$
$$\vec{N} = -(p^4+m)^{-1}\vec{p}\wedge\vec{S} + i\{p^4\nabla_{\vec{p}} + \vec{C}\},$$

where $\vec{C}$ is the 3-vector of components,

$$C^j = (p^4+m)^{-1}\delta^{kl}\left\{(p^ry^r_k + p^sy^s_k)\frac{\partial}{\partial y^j_l} - y^j_k\times\left[p^r\left(\frac{\partial}{\partial y^r_l}\right) + p^s\left(\frac{\partial}{\partial y^s_l}\right)\right]\right\}$$

when $1\leqslant j, r, s\leqslant 3$, and j,r,s are all different.

[1] M. Gell-Mann, Phys. Lett. **8**, 214 (1964); D.W. Greenberg, Phys. Rev. Lett. **13**, 598 (1964); R.H. Dalitz, in *Proceedings of the 13th International Conference on High Energy Physics, Berkeley* 1966 (University of California Press, Berkeley, 1967), and in *Lectures at the 2nd Hawaii Topical Conference on Particle Physics* (University of Hawaii Press, Honolulu, 1967); D. Faiman and A.W. Hendry, Phys. Rev. **173**, 1720 (1968); **180**, 1572 (1969); L.A. Copley, G. Karl, and E. Obryk. Phys. Lett. **29 B**, 117 (1969); D.L. Katyal and A.N. Mitra Phys. Rev. D **1** 338 (1970); D.K. Choudhury and A.N. Mitra. Phys. Rev. D **1**, 351 (1970).

[2] R.P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971); J.L. Rosner, Phys. Rev. D **6**, 1781 (1972); M. Böhm, H. Joos, and M. Krammer, Nucl. Phys. B **51**, 397 (1973) and in *Recent Developments in Mathematical Physics*, P. Urban (Ed.) (Springer-Verlag, New York, 1973); P.J. Walters, A.M. Thomson, and F.D. Gault, J. Phys. A **7**, 1681 (1974).

[3] M. Guenin, in *Mathematical Physics and Physical Mathematics*, K. Maurin and R. Raczka (Eds.) (Reidel, Dordrecht-Holland, 1976).

[4] D. Beau and S. Horchani, C.R. Acad. Sci. (Paris) A **281**, 115 (1975); D. Beau, C.R. Acad. Sci. (Paris) A **281**, 183 (1975).

[5] J. Dixmier, *Algebres Enveloppantes* (Gauthier-Villars, Paris, 1974).

[6] J. Kogut and L. Susskind, Phys. Rev. C **8**, 75 (1973).

[7] I.M. Gelfand and A.A. Kirillov, Inst. Hautes Etudes Sci. Publ. Math. **31**, 5 (1966).

[8] L.P. Horwitz and C. Piron, Helv. Phys. Acta **46**, 316 (1973); L.P. Horwitz, C. Piron, and F. Reuse, Helv. Phys. Acta **48**, 546 (1975); C. Piron and F. Reuse, University of Geneva preprint DPT 1977/07–138.

[9] M. Flato and D. Sternheimer, J. Math. Phys. **7**, 1932 (1966).

[10] S. Horchani, C.R. Acad. Sci. (Paris) A **277**, 201 (1973).

[11] D. Sternheimer, J. Math. Pures Appl. **47**, 247 (1968).

[12] G.W. Mackey, *The Theory of Unitary Group Representations* (University of Chicago Press, Chicago, 1976).

[13] A.S. Wightman, in *Relations de dispersions et particules élémentaires*, C. De Witt and R. Omnes (Eds.) (Hermann, Paris, 1960).

[14] M. Flato and A. Piard, J. Math. Phys. **15**, 1288 (1974).

[15] L. Nachbin, *The Haar Integral* (Van Nostrand, New York, 1965).

[16] *Proceedings of the Seminar on High-energy Physics and Elementary Particles, Trieste* 3 *May*-30 *June* 1965 (IAEA, Vienna, 1965).

[17] M. Flato, in *Mathematical Physics and Physical Mathematics*, K. Maurin and R. Raczka (Eds.) (Reidel, Dordrecht-Holland, 1976).

[18] M. Flato and D. Sternheimer, Phys. Rev. Lett. **16**, 1185 (1966); Commun. Math. Phys. **12**, 296 (1969); D. Sternheimer, "Representations locales d'algébres de Lie" (Mimeographed Memoir, Collége de France, 1968).

[19] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1953), Vol. I.

[20] H. Snellman, J. Math. Phys. **19**, 2278 (1978).

[21] G.E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

[22] J.W.B. Hugues, J. Phys. A **6**, 281 (1973).

[23] N. Barash-Schmidt *et al.*, Rev. Mod. Phys. **48**, N° 2, Part II, April 1976; T.G. Trippe *et al.*, Phys. Lett. B **68**, 1 (1977).

[24] E. Angelopoulos, Ann. Inst. Henri Poincaré **18**, 39 (1973).

[25] M. Flato and H. Snellman, J. Math. Phys. **15**, 857 (1974).

A uniqueness result in the Segal–Weinless approach to linear Bose fields

Bernard S. Kay[a)]
Institut Für Theoretische Physik, Universität Bern, Sidlerstrasse 5, 3012 Berne, Switzerland

(Received 19 December 1978)

We prove a theorem, which, while it fits naturally into the Segal–Weinless approach to quantization seems to have been overlooked in the literature: Let (D,σ) be a symplectic space, and $\mathscr{T}(t)$ a one parameter group of symplectics on (D,σ). Let (H, $2\mathrm{Im}\langle\cdot|\cdot\rangle$) be a complex Hilbert space considered as a real symplectic space, and $\mathrm{U}(t)$ a one-parameter unitary group on H with strictly positive energy. Suppose there is a linear symplectic map K from D to H with dense range, intertwining $\mathscr{T}(t)$ and $\mathrm{U}(t)$. Then K is unique up to unitary equivalence.

I. INTRODUCTION

In the Segal–Weinless approach to linear Bose fields,[1] one is concerned with giving a mathematically rigorous account of the so-called "second-quantization" process for linear field theories.

To model the classical dynamics of a (possibly infinite dimensional) linear system, we adopt Weinless' definition:

Definition 1[2]: *A boson single particle space* $(D,\sigma,\mathscr{T}(t))$ consists of a real symplectic space (D,σ) together with a one-parameter symplectic group $\mathscr{T}(t)$, i.e.,

$$\sigma(\mathscr{T}(t)\Phi,\mathscr{T}(t)\Psi) = \sigma(\Phi,\Psi) \quad \forall\ \Phi,\Psi\in D.$$

Important examples of boson single particle spaces are ones of the form $(\mathscr{H}, 2\mathrm{Im}\langle\cdot|\cdot\rangle, \mathrm{U}(t))$, where $\mathscr{H}$ is some complex Hilbert space, considered as a real vector space, $\langle\cdot|\cdot\rangle$ is the usual Hilbert space inner product, and $\mathrm{U}(t)$ a unitary group on $\mathscr{H}$ with strictly positive energy:

Definition 2[2]: A unitary group is said to have *strictly positive energy* if

(a) $\mathrm{U}(t)$ is weakly continuous, thus $\mathrm{U}(t) = e^{-iHt}$, and

(b) The self-adjoint generator H is positive with dense range. These examples may be thought of as "first-quantized" systems. For such "first-quantized" systems, there exists a straightforward quantization procedure[3] often known as "second quantization." Suffice it to say here that this straightforward quantization is characterized by the "generating functional" or "vacuum state"

$$\langle\Omega\,|\,W(x)\Omega\rangle = \exp(-\|x\|^2/2), \quad x\in\mathscr{H},$$

where $W(x)$ is the Weyl operator, Ω the vacuum. (A precise definition of quantization is given for example by Weinless who defines a "positive-energy Bose–Einstein field"-see §4 of Ref. 2.)

To quantize an arbitrary boson single particle space, one then adopts the following strategy: First, one seeks a "single particle structure":

Definition 3: A single particle structure $(K,\mathscr{H},\mathrm{U}(t))$ for a single particle space $(D,\sigma,\mathscr{T}(t))$ consists of a complex Hilbert space $\mathscr{H}$, a unitary group with strictly positive energy $\mathrm{U}(t)$ and a real linear map K from D to $\mathscr{H}$ satisfying:

(1) ran$K$ is dense in $\mathscr{H}$

(2) $2\mathrm{Im}\langle K\Phi\,|\,K\Psi\rangle = \sigma(\Phi,\Psi)$ (i.e., K is symplectic),

(3) $K(\mathscr{T}(t)\Phi) = \mathrm{U}(t)K(\Phi)$ [i.e., K intertwines $\mathscr{T}(t)$ and $\mathrm{U}(t)$].

Roughly speaking then, one seeks to simultaneously "Hilbertize" D and "unitarize" $\mathscr{T}(t)$, enabling us to quantize by the method of second quantization. More precisely, we can then take the quantization for our single particle space determined by the generating functional:

$$\langle\Omega\,|\,W(\Phi)\Omega\rangle = \exp(-\|K(\Phi)\|^2/2), \quad \Phi\in D.$$

The question arises as to the existence and uniqueness of single particle structures. The existence question recently acquired renewed interest in the currently popular subject of "quantum field theory in curved space–times." And, in a recent paper,[4] we proved an existence theorem for a special class of single particle spaces arising when one quantizes the covariant Klein–Gordon system on a stationary space–time.

The purpose of this paper is to show that, when a single particle structure exists, it is unique. An analogous result for Fermi–Dirac fields has already been given in the Weinless paper,[2] but the boson case seems to have been overlooked there (though the result is implicitly assumed).

For the perhaps more important question of quantization itself, the reader is referred again to Segal[1] and Weinless[2] (see also Ref. 5) where results are given on the uniqueness (and lack of uniqueness!) of quantization.

II. A UNIQUENESS THEOREM

Theorem: Suppose there exists a single particle structure $(K,\mathscr{H},\mathrm{U}(t))$ for a single particle space $(D,\sigma,\mathscr{T}(t))$; then it is unique up to unitary equivalence.

Proof: Suppose there are two such single particle structures $(K_1,\mathscr{H}_1,\mathrm{U}_1(t))$, $(K_2,\mathscr{H}_2,\mathrm{U}_2(t))$: Then $T := K_2\circ K_1^{-1}$ is a real, linear invertible, symplectic from the real linear dense domain $K_1(D)$ in H_1 to the real linear dense domain $K_2(D)$ in H_2. Also (3) of definition 3 implies $K_1(D)$ is invariant for $U_1(t)$, and $K_2(D)$ is invariant for $\mathrm{U}_2(t)$.

We shall show that T extends to a unitary. We have (a) $T\mathrm{U}_1(t) = \mathrm{U}_2(t)\,T$ on $K_1(D)$, (b) $T^+ \supset -iT^{-1}i$ on domain $iK_2(D)$ where T^+ denotes the real adjoint of T, $H_{1,2}$ considered as real Hilbert spaces (i.e., with inner product $\mathrm{Re}\langle\cdot|\cdot\rangle$). [(b) easily follows since T symplectic.]

a)Work supported by the Schweizerische Nationalfonds.

 0022-2488/79/081712-02$01.00

We now show that T has a closed complex linear extension: Consider

$$f_{x,y}: t \to (\langle U_1(t)x | T^+ y\rangle_{\mathscr{H}_1} - \langle U_2(t)Tx | y\rangle_{\mathscr{H}_2})$$

$$\forall x \in K_1(D);\ y \in D(T^+).$$

By a well known argument (see e.g., Ref. 2) the strictly positive energy of $U_1(t)$, $U_2(t)$ guarantees that $f_{x,y}$ extends to a function which is bounded and holomorphic in the lower half t plane, vanishes as $\mathrm{Im}t \to -\infty$ and is continuous and bounded on the real axis. We also have $\mathrm{Re}f_{x,y}(t)$ vanishes on the real axis by definition of the real adjoint. The Schwartz reflection principle then implies that $f_{x,y}(t)$ is bounded and holomorphic in the plane (and vanishes at ∞). So by Liouville's theorem it vanishes everywhere. We then have, setting $t = 0$,

$$\langle x | T^+ y\rangle_{\mathscr{H}_1} = \langle Tx | y\rangle_{\mathscr{H}_2} \quad \forall x \in K_1(D);\ y \in D(T^+),$$

whereupon

$$\langle Tx | iy\rangle = i\langle Tx | y\rangle = i\langle x | T^{+\,+} y\rangle = \langle x | iT^+ y\rangle,$$

showing $iy \in D(T^+)$ and $T^+(iy) = iT^+ y$. Therefore, T^+ is complex linear, whereupon $\overline{T} = T^{+\,+}$ is complex linear. Taking the closure of both sides of (b), we recover

$$\overline{T}^* \supset -i\overline{T}^{-1} i = \overline{T}^{-1},$$

since, for the complex linear $\overline{T}$, the real adjoint $\overline{T}^+$ and complex adjoint $\overline{T}^*$ coincide. Whereupon $\overline{T}$ is a closed norm- preserving map with dense range, hence unitary. □

[1] I.E. Segal, "Foundations of the Theory of Dynamical Systems of Infinitely Many Degrees of Freedom," I: Mat.-Fys. Medd. Dansk. Vid. Selsk. **31**, No. 12, 1 (1959); (II): Canadian J. Math. **13**, 1 (1961); III: Illinois J. Math. **6**, 500 (1962); I.E. Segal, *Mathematical Problems of Relativistic Physics* (Am. Math. Soc., Rhode Island, Providence, 1963), and Ref. 2.

[2] M. Weinless, "Existence and Uniqueness of the Vacuum for Linear Quantized Fields," J. Funct. Anal. **4**, 350 (1969).

[3] J.M. Cook, "The Mathematics of Second Quantization," Trans. Am. Math. Soc. **21**, 222 (1953).

[4] B.S. Kay, "Linear Spin-Zero Quantum Fields in External Gravitational and Scalar Fields I: A One Particle Structure for the Stationary Case," Commun. Math. Phys. **62**, 55 (1978).

[5] H.J. Borchers, R. Haag, and B. Schroer, "The Vacuum State in Quantum Field Theory," Nuovo Cimento **29**, 148 (1963).

On Blume's integration of Schrödinger's equation for a quantum system subject to random pulses

Henryk Gzyl

Escuela de Física y Matemáticas, Facultad de Ciencias, U.C.V., Caracas, Venezuela
(Received 23 October 1978)

In this note we dispose of Blume's objections to a recent article by Gzyl. We show that he is forgetting about causality and that he makes obvious mistakes taking limits.

In Ref. 1, Blume claims that he shows why the computations in Ref. 2 are wrong. That is, he claims, first, that the equation

$$id\psi_t = H\psi_t\,dt + dV_t\psi_{t-}, \tag{1}$$

where $\psi_t \equiv \psi(t)$ and $\psi_{t-} = \lim_{s\uparrow t}\psi(s)$ and $dV_t = \Sigma_{k=1}^{\infty} V_k \times \delta(t - T_k)\,dt$, which models a quantum system to random pulses, is incorrect and, second, that it is incorrectly integrated.

Blume claims that instead of (1) one should consider

$$id\psi_t = H\psi_t\,dt + dV_t\psi_t. \tag{2}$$

Here we show why one must consider (1) instead of (2) and that even if (2) were the choice, Blume's computations are totally wrong. We do this by showing that (1) is causally correct, and we explicitly carry out the limiting procedures in Ref. 1 and show where he goes wrong.

Notice first that between pulses the system evolves according to the unperturbed Hamiltonian H, and therefore in order to integrate (1) [or (2)] it is enough to consider a quantum system subject to just one pulse at a fixed time T. Thus, instead of (1) or (2) consider the equations

$$id\psi_t = H\psi_t\,dt + V\delta(t - T)\psi_{t-}\,dt, \tag{3}$$

$$id\psi_t = H\psi_t\,dt + V\delta(t - T)\psi_t\,dt, \tag{4}$$

which can be rewritten as

$$d\psi_t + iH\psi_t\,dt = -iV\delta(t - T)\psi_{t-}\,dt,$$

$$d\psi_t + iH\psi_t\,dt = -iV\delta(t - T)\psi_t\,dt.$$

Now multiply both sides of each equation by e^{iHt} and integrate from 0 to t to obtain

$$e^{iHt}\psi(t) - \psi(0) = -i\int_0^t e^{iHs}V\delta(s - T)\psi_{s-}\,ds, \tag{5}$$

$$e^{iHt}\psi(t) - \psi(0) = -i\int_0^t e^{iHs}V\delta(s - T)\psi_s\,ds, \tag{6}$$

Now, if $t < T$, both equations yield that $\psi(t) = e^{-iHt}\psi(0)$, since up to T there has been no pulse on the system. Note also that $\psi(T-) = \lim_{s\uparrow T}\psi(s) = e^{-iHT}\psi(0)$. Consider now the case $t = T$, then Eqs. (5) and (6) yield

$$\psi(T) - \psi(T-) = -iV\psi(T-), \tag{7}$$

$$\psi(T) - \psi(T-) - iV\psi(T), \tag{8}$$

respectively. Comparing (7) and (8), we see that in (7) the size of the discontinuity produced by the pulse at T depends on $\psi(T-)$, the state of the system "right before T," whereas in (8) the size of the discontinuity depends on $\psi(T)$, the state of the system "right after T."

Therefore, if we want our model to be causal, we are led to choose (1) instead of (2) for describing the time evolution of the system.

But suppose we did not care about causality, or that we are forced to consider (2) [or (6)] instead to (1) [or (5)]. As noted above, from (6) we obtain (8) from which we obtain $\psi(T) = (1 + iV)^{-1}\psi(T-) = (1 + iV)^{-1}e^{-iHT}\psi(0)$, and from this it follows that for $t \geqslant T$ the state of the system is given by, $\psi(t) = e^{-iH(t-T)}(1 + iV)^{-1}e^{-iHT}\psi(0)$.

Had we used (5) insteads of (6), a similar computation would have lead us to

$$\psi(t) = e^{-iH(t-T)}(1 - iV)e^{-iHT}\psi(0). \tag{9}$$

Besides the fact (missed by Blume) that (6) can be trivially integrated, it is easy to see from either (7) or (8) that probability is not conserved at the time of the pulses. Again, this is due to the fact that the system is under a time dependent perturbation, and therefore not isolated, or it is a defect of the model. We emphasize that the nonconservation of probability has nothing to do either with randomness or with the singular nature of the perturbation.

Let us examine now the procedure which Blume calls "the correct handling of the singularity" (which could have been avoided as we did show above). Rewrite (6) as $d\phi_t = -i\hat{V}(t)\delta(t - T)\phi_t\,dt$, where $\phi_t = e^{iHt}\psi_t$, $\hat{V}(t) = e^{-iHt}V^{-iHt}$, which corresponds to (4) in Ref. 1. Suppose, furthermore, to go along with Ref. 1, that V and H commute; then (2) becomes $d\phi = -iV\delta(t - T)\phi_t\,dt$. Instead of integrating this equation, Blume prefers to do things the hard way and considers

$$d\phi_\epsilon = -iVf(t)\phi_\epsilon(t)\,dt \tag{10}$$

where $\epsilon f(t) = I_{(T+\epsilon/2,\,T+\epsilon/2)}(t)$, the indicator function of the interval $(T - \epsilon/2,\ T + \epsilon/2)$. We put the ϵ as subscript in $\phi_\epsilon(t)$ to indicate its dependence on ϵ. Then Blume says (and this happens to be right!) that $\phi_\epsilon(t) = \exp[-iV\int_0^t f(s)\,ds]\psi(0)$. Now, notice that

$$\phi_\psi(t) = \begin{cases} \psi_0 & \text{if } t \leqslant T - \epsilon/2, \\ \exp[-iV(t - T + \epsilon/2)/\epsilon]_{\psi_0} & \text{if } T - \epsilon/2 \leqslant t \leqslant T + \epsilon/2, \\ \exp(-iV)\psi_0 & \text{if } T + \epsilon/2 < t, \end{cases}$$

from which anybody can see that $\phi_\epsilon(T) = \exp(-iV/2)\psi_0$,

 0022-2488/79/081714-02$01.00

and that if we let $\epsilon \to 0$, we obtain

$$\lim_{\epsilon \to 0} \phi_\epsilon(t) = \begin{cases} \psi_0, & t < T, \\ \exp(-iV/2)\psi_0, & t = T, \\ \exp(-iV)\psi_0, & T < t, \end{cases}$$

which by no means tends to $e^{iHt}\psi(t)$, with $\psi(t)$ as above, as it should.

This is another instance of a singular perturbation problem: the solution of (2), considered as a limit of (10) as $\epsilon \to 0$ in some appropriate sense, is different from the limit of the solutions of (10).

Note: The technique for integrating equations like (1) is nothing new. For details and references see Ref. 3.

[1] M. Blume, J. Math. Phys. **19**, 2004 (1978).
[2] H. Gzyl, J. Math. Phys. **18**, 1327 (1977).
[3] D. Snyder, *Random Point Processes* (Wiley, New York, 1975).

Settling the question of the high-energy behavior of phase shifts produced by repulsive, strongly singular, inverse-power potentials

Nanny Fröman and Karl-Erik Thylwe
Institute of Theoretical Physics, University of Uppsala, Uppsala, Sweden

(Received 20 November 1978)

For repulsive, strongly singular, inverse-power potentials it is rigorously shown that the JWKB expression for the nonrelativistic phase shift tends to exactness in the high-energy limit. The hitherto open question as to the correct expression for the leading term in the high-energy expansion of the phase shift for these potentials is thus definitely settled, and it is further confirmed that even the next term in the expansion yielded by the JWKB expression is significant.

1. INTRODUCTION

The high-energy limit of the nonrelativistic scattering phase-shift produced by strongly singular potentials has been studied analytically by several authors. For a general review, the reader is referred to the extensive article by Frank *et al.* [1] While there is agreement as to the form of the energy dependence of the leading term, the expressions for the coefficient of this term, obtained by different methods, do not agree, though they give similar numerical values (see Table I in Ref. 1). Thus, for pure inverse-power potentials, i.e., for $V(r)$, in units of $\hbar^2/(2m)$, given by

$$V(r) = g^2 r^{-n}, \quad n > 2, \tag{1.1}$$

where g^2 is a coupling constant, various authors [2–7] agree that

$$\delta_l(k) \underset{k \to \infty}{\sim} -A_{n,0}\chi, \tag{1.2}$$

where

$$\chi = k(g^2/k^2)^{1/n} = g^{2/n}k^{(n-2)/n} \to \infty, \quad \text{when } k \to \infty, \tag{1.3}$$

but for the constant $A_{n,0}$ the following differing expressions are given:

$$A_{n,0} = \frac{2^{1-2/n}\pi/n}{\sin(\pi/n)\Gamma(3-2/n)}, \quad \text{Calogero [2]}, \tag{1.4}$$

$$A_{n,0} = \frac{\sqrt{\pi}}{2}\frac{\Gamma(1-1/n)}{\Gamma(3/2-1/n)}, \quad \text{Bertocchi } et\ al.,\ [5]$$

$$\text{Paliov and Rosendorff,}\ [6] \tag{1.5}$$

$$A_{n,0} = \frac{1-1/(2n)}{1-1/n}, \quad \text{Jabbur.}\ [7] \tag{1.6}$$

Calogero obtained his result by a variable-phase calculation, Bertocchi *et al.* as well as Paliov and Rosendorff used the JWKB method, and Jabbur worked with Volterra integral equations, matching solutions valid for small and large r, respectively, at the classical turning point. In a recent work, suggested by Calogero, Dolinszky [8] has numerically investigated the scattering problem for a potential proportional to $1/r^4$. His results support, for that particular potential, the asymptotic exactness of the result obtained by means of the JWKB approximation, i.e., (1.5). However, a decisive and general proof, valid for all $n > 2$, is lacking hitherto.

In the present paper we shall show rigorously that in the high-energy limit the JWKB approximation yields the phase shift, produced by the class of potentials (1.1), with an error proportional to χ^{-1}, which, in turn, implies that the JWKB expression for the phase-shift yields correctly not only the term proportional to χ but also the constant term in the high-energy expansion of $\delta_l(k)$.

Our proof of the above-mentioned result is based on the theory for mastering connection problems developed in Ref. 9. We quote in Sec. 2 below an exact formula for the phase shift derived in Ref. 9 and confirm that the conditions for the validity of this formula are fulfilled also for the case of strongly singular potentials. This exact formula gives the phase shift as the JWKB expression plus a correction term for which an upper bound is obtainable. In Sec. 3 we show that, for strongly singular potentials of the form (1.1), the correction term tends to zero at least as fast as χ^{-1}, when the energy tends to infinity.

2. EXACT FORMULA FOR THE PHASE-SHIFT

We write the radial Schrödinger equation as

$$\frac{d^2u}{dr^2} + Q^2(r)u = 0, \tag{2.1}$$

where

$$Q^2(r) = k^2 - V(r) - l(l+1)/r^2 \tag{2.2}$$

with obvious notations. The wavenumber k and the orbital angular momentum quantum number l are real ($k > 0$, $l \geqslant 0$), whereas r is allowed to assume complex values.

For strongly singular potentials it is not necessary to replace $l(l+1)$ by $(l+\frac{1}{2})^2$ in the JWKB functions, the latter being good at the origin whether we modify in this way or not (see p. 114 in Ref. 9, where a criterion is given; cf. also p. 40 in Ref. 1). It is, however, convenient to use this modification, since the JWKB formula for the phase shift is then exact in the limit $g^2 = 0$. Thus, we choose

$$Q^2_{\text{mod}}(r) = k^2 - V(r) - (l+\tfrac{1}{2})^2/r^2. \tag{2.3}$$

 0022-2488/79/081716-04$01.00

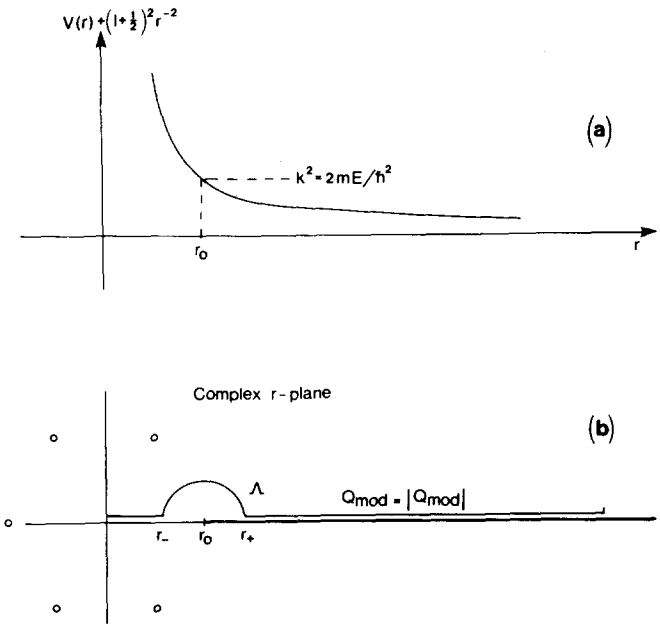

FIG. 1. (a) Qualitative behavior of the effective potential $V(r) + (l+\frac{1}{2})^2 r^{-2}$. (b) Path of integration Λ for the μ integral (3.2). The heavy line from r_0 towards $r = +\infty$ indicates a cut. The phase of $Q_{\rm mod}(r)$ on the upper edge of the cut is indicated. The small circles show the approximate location of further zeros of $Q^2_{\rm mod}(r)$, beside r_0, for the case $n = 6$.

The exact phase-shift formula (11.40b) in Ref. 9, on which our proof will be based, was derived for the case that the physical potential is at the most as singular as r^{-2} at the origin. The formula is, however, valid also for strongly singular potentials, since the two decisive conditions are fulfilled, namely that the quantity $\exp[\int^r |Q_{\rm mod}(r)dr|]$ increases monotonically and tends to infinity as $r \to +0$ along the real axis and that, furthermore, the integral $\int^r |\epsilon Q_{\rm mod} dr|$, where

$$\epsilon = \frac{Q^2 - Q^2_{\rm mod}}{Q^2_{\rm mod}} + \frac{1}{16 Q^6_{\rm mod}} \times \left[5\left(\frac{dQ^2_{\rm mod}}{dr}\right)^2 - 4Q^2_{\rm mod}\frac{d^2 Q^2_{\rm mod}}{dr^2}\right], \tag{2.4}$$

is convergent when $r \to +0$ along the real axis, as is seen when one notices that, for $n > 2$, the integrand, $|\epsilon Q_{\rm mod}|$, behaves as $r^{n/2-2}$ in the immediate neighborhood of the origin. Thus, in accordance with (11.40b) in Ref. 9, we have the exact formula

$$\delta_l(k) = [\delta_l(k)]_{\rm JWKB} - \arg[F_{11}(+0, +\infty)], \tag{2.5}$$

where $-\arg[F_{11}(+0, +\infty)]$ is the correction term which we shall examine in the next section, and

$$[\delta_l(k)]_{\rm JWKB} = \lim_{R \to +\infty}\left(\int_{r_0}^{R} Q_{\rm mod}(r)dr - \int_{(l+1/2)/k}^{R} [k^2 - (l+\tfrac{1}{2})^2/r^2]^{1/2} dr\right), \tag{2.6}$$

r_0 being the generalized classical turning point, i.e., the zero of $Q^2_{\rm mod}(r)$ on the real axis (see Fig. 1).

3. ESTIMATE OF ARG $F_{11}(+0, +\infty)$

According to the estimate given on p. 128 in Ref. 9, we have

$$|\arg[F_{11}(+0, +\infty)]| \leqslant \mu + \text{higher powers of } \mu, \tag{3.1}$$

where

$$\mu = \int_\Lambda |\epsilon Q_{\rm mod} dr|, \tag{3.2}$$

the integration being performed along a path Λ from $+0$ to $+\infty$, on which there is only one extremum of $|\exp(i\int' Q_{\rm mod} dr)|$. See Sec. 6.3 and Fig. 6.1 in Ref. 9.

In order to estimate the μ integral (3.2), we shall start by considering the explicit expression for ϵ. With $V(r)$ given by (1.1) and $Q^2(r)$ and $Q^2_{\rm mod}(r)$ defined by (2.2) and (2.3), respectively, we readily obtain from (2.4)

$$\epsilon Q_{\rm mod} = \frac{1}{4r^2 Q_{\rm mod}} + \frac{1}{4Q^3_{\rm mod}}\left(\frac{n(n+1)g^2}{r^{n+2}} + \frac{6(l+\frac{1}{2})^2}{r^4}\right) + \frac{5}{16 Q^5_{\rm mod}}\left(\frac{n^2 g^4}{r^{2n+2}} + \frac{4ng^2(l+\frac{1}{2})^2}{r^{n+4}} + \frac{4(l+\frac{1}{2})^4}{r^6}\right). \tag{3.3}$$

This relation can be written

$$\epsilon Q_{\rm mod} = \sum_{i=1}^{3}\sum_{j=0}^{i-1} C_{ij} \frac{g^{2j} r^{(n-2)(i-j-1/2)-1}}{[r^n Q^2_{\rm mod}(r)]^{i-1/2}}, \tag{3.3a}$$

or, alternatively,

$$\epsilon Q_{\rm mod} = \sum_{i=1}^{3}\sum_{j=0}^{i-1} C_{ij} \frac{g^{2j} r^{-2i-(n-2)j}}{[Q^2_{\rm mod}(r)]^{i-1/2}}, \tag{3.3b}$$

where the quantities C_{ij}, as is easily realized, are independent of g, k, and r, but may possibly depend on n and l.

In order to be able to choose the path of integration Λ conveniently, we must know the location of the transition points, i.e., the roots of the equation

$$Q^2_{\rm mod}(r) = 0. \tag{3.4}$$

With the aid of (2.3), (1.1), and (1.3), we can write (3.4) as

$$\left[r\left(\frac{k^2}{g^2}\right)^{1/n}\right]^n - \frac{(l+\frac{1}{2})^2}{\chi^2}\left[r\left(\frac{k^2}{g^2}\right)^{1/n}\right]^{n-2} = 1. \tag{3.5}$$

Recalling that $\chi \to \infty$ in the high-energy limit [see (1.3)], we realize that the roots r_m of Eq. (3.5) are

$$r_m \underset{k\to\infty}{=} (g^2/k^2)^{1/n} e^{i2\pi m/n}[1 + O(\chi^{-2})], \quad m = 0, \ldots, n-1. \tag{3.6}$$

The distance between any two neighboring roots r_m and $r_{m'}$, is given by

$$|r_m - r_{m'}| \underset{k\to\infty}{=} 2(g^2/k^2)^{1/n}\sin(\pi/n)[1 + O(\chi^{-2})]. \tag{3.7}$$

The generalized classical turning point r_0 is obtained by putting $m = 0$ in (3.6).

The path Λ shall circumvent the turning point r_0, at which the integral of $|\epsilon Q_{\rm mod}|$ diverges. In Fig. 1b such a path Λ is depicted, chosen so as to proceed along the real axis and to circumvent r_0 in a semicircle enclosing no other zeros of $Q^2_{\rm mod}(r)$. Since $r_0 \to 0$ when $k \to \infty$, we choose conveniently

the radius of the semicircle as a definite fraction η (independent of k) of the distance r_0. Furthermore, by choosing η small enough, we achieve the result that $Q^2_{\mathrm{mod}}(r)$ is approximately a linear function of $r - r_0$ on the semicircle (see the Appendix), and we thereby assure the fulfillment of the above-mentioned condition that there shall be only one extremum (with our choice of phase of $Q_{\mathrm{mod}}(r)$ a minimum) of $|\exp(i\int^r Q_{\mathrm{mod}} dr)|$ on the path Λ.

To obtain an upper bound for the μ integral (3.2), we write

$$\mu = \mu_1 + \mu_2 + \mu_3, \tag{3.8}$$

where μ_1 and μ_3 are the integrals along the two parts of Λ coinciding with the real axis and μ_2 is the integral along the semicircle joining them (see Fig. 1b). We shall now examine these integrals separately.

Let us first estimate μ_1. To this purpose we consider the integral over the absolute value of an arbitrary term in the right-hand member of (3.3a) from $r = +0$ to $r = r_-$, where, with due regard to (3.6), we have

$$r_- = (1-\eta) r_0 \underset{k\to\infty}{=} (1-\eta)(g^2/k^2)^{1/n}[1 + O(\chi^{-2})]. \tag{3.9}$$

The region from the origin to the classical turning point is classically forbidden, and hence $Q^2_{\mathrm{mod}}(r) \leqslant 0$ for $0 \leqslant r \leqslant r_0$. According to (2.3) and (1.1), we therefore have

$$|r^n Q^2_{\mathrm{mod}}(r)| = g^2 - k^2 r^n + (l+\tfrac{1}{2})^2 r^{n-2}, \quad 0 \leqslant r \leqslant r_0. \tag{3.10}$$

Putting $r = r_-$ in (3.10) and using (3.9), we obtain in the high-energy limit the relation

$$|r_-^n Q^2_{\mathrm{mod}}(r_-)| \underset{k\to\infty}{\sim} g^2[1 - (1-\eta)^n], \tag{3.11}$$

which will be used presently. Examining, furthermore, the behavior of the function $|r^n Q^2_{\mathrm{mod}}(r)|$ in the interval $(+0, r_-)$, we find from (3.10) that, for sufficiently large values of k, it takes its smallest value for $r = r_-$. Utilizing this fact and taking (3.11) and (3.6) into account, we obtain the following upper bound for the integral over the absolute value of an arbitrary term in the right-hand member of (3.3a) (suppressing the factor C_{ij} which is immaterial here):

$$\begin{aligned}
&\int_0^r \frac{g^{2j} r^{(n-2)(i-j-1/2)-1}}{|r^n Q^2_{\mathrm{mod}}(r)|^{i-1/2}}\, dr \\
&\leqslant \frac{g^{2j}}{|r_-^n Q^2_{\mathrm{mod}}(r_-)|^{i-1/2}} \int_0^r r^{(n-2)(i-j-1/2)-1} dr \\
&< \frac{g^{2j}}{|r_-^n Q^2_{\mathrm{mod}}(r_-)|^{i-1/2}} \int_0^{r_0} r^{(n-2)(i-j-1/2)-1} dr \\
&= \frac{g^{2j}}{|r_-^n Q^2_{\mathrm{mod}}(r_-)|^{i-1/2}} \\
&\quad \times [(n-2)(i-j-\tfrac{1}{2})]^{-1} r_0^{(n-2)(i-j-1/2)} \\
&\underset{k\to\infty}{\sim} [1-(1-\eta)^n]^{-i+1/2} \\
&\quad \times \psi[(n-2)(i-j-\tfrac{1}{2})]^{-1} \chi^{1-2(i-j)} \\
&= O(\chi^{1-2(i-j)}).
\end{aligned} \tag{3.12}$$

Since $i - j \geqslant 1$ in (3.3a), we realize that the slowest approach to zero of the integral (3.12), when $k \to \infty$, is exhibited when $i - j$ takes its smallest value, unity. We thus conclude that $\mu_1 = O(\chi^{-1})$, when $k \to \infty$.

An upper bound for μ_3 is obtained analogously as for μ_1 by utilizing the fact that, in the interval $(r_+, +\infty)$ with

$$r_+ = (1+\eta) r_0, \tag{3.13}$$

the function $Q^2_{\mathrm{mod}}(r)$ is a monotonically increasing function of r. Thus, in view of the fact that $i \geqslant 1$ and with the aid of the expressions (3.6) and (3.13) for r_0 and r_+, respectively, we get

$$\begin{aligned}
\frac{1}{|Q^2_{\mathrm{mod}}(r)|^{i-1/2}} &\leqslant \frac{1}{|Q^2_{\mathrm{mod}}(r_+)|^{i-1/2}} \\
&\underset{k\to\infty}{\sim} [1-(1+\eta)^{-n}]^{-i+1/2} k^{-2i+1}.
\end{aligned} \tag{3.14}$$

Estimating the integral over the absolute value of an arbitrary term in the right-hand member of (3.3b) from $r = r_+$ to $r = +\infty$, we thus obtain (suppressing C_{ij})

$$\begin{aligned}
&\int_{r_+}^\infty \frac{g^{2j} r^{-2i-(n-2)j}}{|Q^2_{\mathrm{mod}}(r)|^{i-1/2}}\, dr \\
&\leqslant \frac{g^{2j}}{|Q^2_{\mathrm{mod}}(r_+)|^{i-1/2}} \int_{r_+}^\infty r^{-2i-(n-2)j} dr \\
&< \frac{g^{2j}}{|Q^2_{\mathrm{mod}}(r_+)|^{i-1/2}} \int_{r_0}^\infty r^{-2i-(n-2)j} dr \\
&\underset{k\to\infty}{\sim} [1-(1+\eta)^{-n}]^{-i+1/2} \\
&\quad \times [2i + (n-2)j - 1]^{-1} \chi^{1-2(i-j)} \\
&= O(\chi^{1-2(i-j)}).
\end{aligned} \tag{3.15}$$

This upper bound displays the same high-energy behavior as the upper bound obtained in (3.12). Hence, also $\mu_3 = O(\chi^{-1})$ when $k \to \infty$.

Finally an upper bound for the integral over the absolute value of an arbitrary term in (3.3b) along the semicircle joining the points r_- and r_+ is obtained, simply by multiplying the maximum value of the integrand by the length of the path. According to our previous assumption the radius of the semicircle is ηr_0, and on the semicircle we therefore have

$$|r - r_0| = \eta r_0. \tag{3.16}$$

If we choose the parameter η sufficiently small and, as already assumed, independent of k, we may, in the high-energy limit, use the expression (A5) in the Appendix for the function $Q^2_{\mathrm{mod}}(r)$ on the semicircle. Noting, furthermore, that $|r| \geqslant r_-$ on the semicircle, we obtain the following upper bound for the absolute value of an arbitrary term in (3.3b) (with C_{ij} suppressed):

$$\begin{aligned}
\frac{g^{2j} r^{-2i-(n-2)j}}{|Q^2_{\mathrm{mod}}(r)|^{i-1/2}} &\underset{k\to\infty}{\leqslant} g^{2j} r_-^{-2i-(n-2)j} (nk^2\eta)^{1/2-i} \\
&\quad \times [1 + O(\eta) + O(\chi^{-2})].
\end{aligned} \tag{3.17}$$

On multiplying the right-hand member of (3.17) by the

length of the path, $\pi\eta r_0$, and substituting the expressions (3.6) and (3.9) for r_0 and r_-, respectively, we find that the upper bound for the integral in question equals $O(\chi^{1-2(i-j)})$, when $k\to\infty$. This implies that $\mu_2 = O(\chi^{-1})$ in the high-energy limit.

Thus we conclude that

$$\mu = \mu_1 + \mu_2 + \mu_3 = O(\chi^{-1}). \tag{3.18}$$

Hence, it follows from (3.1) that the correction term $\arg[F_{11}(+0,+\infty)]$ occurring in the exact formula (2.5) tends to zero at least as fast as χ^{-1} when $k\to\infty$. Thus it is rigorously shown that the JWKB expression (2.6) for the phase shift tends to exactness as $k\to\infty$. From (2.6) it then follows that, in the high-energy limit, we have, for fixed l, the expansion

$$\delta_l(k) \underset{k\to\infty}{=} -A_{n,0}\chi + (l+\tfrac{1}{2})\pi/2 + O(\chi^{-1}), \tag{3.19}$$

where $A_{n,0}$ and χ are given by (1.5) and (1.3), respectively.

The question of which one (if any) of the expressions (1.4–1.6) for $A_{n,0}$ is correct is thus definitely settled, the expression (1.5) being the correct one. Also the next term in the high-energy expansion obtained from (2.6), i.e., the constant $(l+\frac{1}{2})\pi/2$ in (3.19), is significant according to our analysis.

We remark that our treatment can easily be generalized to apply to a class of inverse-power potentials with an energy dependent coupling constant. In fact, if we replace g^2 in (1.1) by $g^2 = g'^2 k^a$, where $-(n-2) < a \leqslant 2 < n$, the expansion (3.19), with χ replaced by $\chi' = k(g'^2/k^{2-a})^{1/n}$, still applies also for the class of energy dependent potentials in question.

APPENDIX: LINEAR APPROXIMATION OF THE FUNCTION $Q^2_{\rm mod}(r)$ CLOSE TO THE TURNING POINT r_0

Consider the function $Q^2_{\rm mod}(r)$ in the region of the complex r plane defined by $|r-r_0| \leqslant \eta r_0$, $0<\eta<1$. In this region $Q^2_{\rm mod}(r)$ is an analytical function of r and can thus be expanded in a Taylor expansion about the point r_0. Since $Q^2_{\rm mod}(r_0) = 0$, the Taylor series is

$$\begin{aligned} Q^2_{\rm mod}(r) &= \sum_{\nu=1}^{\infty} \frac{1}{\nu!}\left(\frac{d^\nu}{dr^\nu} Q^2_{\rm mod}(r)\right)_{r=r_0} (r-r_0)^\nu \\ &= k^2 \sum_{\nu=1}^{\infty} a_\nu \left(\frac{r}{r_0}-1\right)^\nu, \end{aligned} \tag{A1}$$

where

$$\begin{aligned} a_\nu &= \frac{1}{\nu! k^2} r_0^\nu \left(\frac{d^\nu}{dr^\nu} Q^2_{\rm mod}(r)\right)_{r=r_0} \\ &= (-1)^{\nu-1}\binom{n-1+\nu}{\nu}\frac{g^2}{k^2 r_0^n} \\ &\quad \times \psi\left(1 + \frac{(n-1)!(\nu+1)!}{(n-1+\nu)!}\frac{(l+\frac{1}{2})^2}{g^2} r_0^{n-2}\right). \end{aligned} \tag{A2}$$

Inserting the expression (3.6) for r_0 into (A2) and using the definition (1.3), we get

$$a_\nu = (-1)^{\nu-1}\binom{n-1+\nu}{\nu}[1 + O(\chi^{-2})], \quad \text{when } k\to\infty, \tag{A3}$$

where $O(\chi^{-2})$ is uniformly bounded with respect to ν. Comparing the absolute values of two consecutive terms in the right-hand member of (A1), we obtain, with the aid of (A3) and the fact that $|r-r_0| \leqslant \eta r_0$ in the region under consideration,

$$\begin{aligned} &\left|\frac{a_{\nu+1}(r/r_0-1)^{\nu+1}}{a_\nu (r/r_0-1)^\nu}\right| \\ &\quad \underset{k\to\infty}{=} \frac{n+\nu}{\nu+1}\frac{|r-r_0|}{r_0}[1+O(\chi^{-2})] \\ &\quad \leqslant \tfrac{1}{2}(n+1)\eta + O(\chi^{-2}), \quad \nu \geqslant 1. \end{aligned} \tag{A4}$$

Thus, if we choose $\eta \ll 2/(n+1)$, the terms in the expansion (A1) will in the high-energy limit decrease rapidly with increasing values of ν. Retaining explicitly only the term which is linear in $r-r_0$, we realize from (A1) and (A3) that

$$Q^2_{\rm mod}(r) \underset{k\to\infty}{=} nk^2(r/r_0-1)[1+O(\eta)+O(\chi^{-2})], \quad \eta \ll 2/(n+1), \tag{A5}$$

in the region $|r-r_0| \leqslant \eta r_0$.

[1] W.M. Frank, D.J. Land, and R.M. Spector, Rev. Mod. Phys. **43**, 36 (1971).
[2] F. Calogero, Phys. Rev. **135**, B693 (1964).
[3] F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York and London, 1967).
[4] L. Bertocchi, S. Fubini, and G. Furlan, Nuovo Cimento **35**, 599 (1965).
[5] L. Bertocchi, S. Fubini, and G. Furlan, Nuovo Cimento **35**, 633 (1965).
[6] A. Paliov and S. Rosendorff, J. Math. Phys. **8**, 1829 (1967).
[7] R.J. Jabbur, Phys. Rev. **138**, B1525 (1965).
[8] T. Dolinszky, Nuovo Cimento A **22**, 578 (1974).
[9] N. Fröman and P.O. Fröman, *JWKB-Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965) (Russian translation: MIR, Moscow, 1967).

Classical, cross-section generating solutions of field equations

Peter Suranyi
Blackett Laboratory, Imperial College, London SW7 2AZ, England

(Received 6 March 1979; accepted for publication 20 April 1979)

The symmetry properties of classical, cross-section generating solutions of field equations, sectons, are investigated. It is shown that under general conditions on the interaction Lagrangian of the field theory the symmetry group of the solutions can only be $0(1,1)\times 0(2)$ for finite momenta. Such solutions generate inclusive cross sections with Feynman scaling.

I. INTRODUCTION

A functional integral representation of the generating functional of cross sections for an arbitrary field theory was found in a previous work.[1] This representation was used to show that under some circumstances, inclusive (and semi-inclusive) cross sections are dominated by certain classical solutions of the field equations, called sectons.[1,2]

The simplest example for the use of the theory of sectons is the calculation of the single particle inclusive cross section that has the following exact form

$$d\sigma/d^3p = \sigma(2\pi)^{-3}(2E)^{-1}\langle j(p)\bar{j}(-p)\rangle_1, \tag{1.1}$$

where $j(p)$ and $\bar{j}(p)$ are the Fourier transforms of $(\Box + m^2)\psi(x)$ and $(\Box + m^2)\bar{\psi}(x)$, respectively; $p_0 = E = (\mathbf{p}^2 + m^2)^{1/2}$ and the expectation value of an operator, $\langle O[\psi,\bar{\psi}]\rangle_g$ is defined through the functional integral

$$\begin{aligned}\langle O[\psi,\bar{\psi}]\rangle_g &= \sigma^{-1}m^2[(p_a\cdot p_b)^2 - m^4]^{-1/2}64\int dz\,dv\,dv' \\ &\times\exp\{i(v-v')\cdot(p_a - p_b) + iz\cdot(p_a + p_b)\} \\ &\times\int D\psi\,D\bar{\psi}O[\psi,\bar{\psi}] \\ &\times K\psi(v + z/2)K\psi(-v + z/2)K\bar{\psi}(v' - z/2) \\ &\times K\bar{\psi}(-v' - z/2)e^{-A[g]},\end{aligned} \tag{1.2}$$

where $K = \Box + m^2$, m is the physical mass, while p_a and p_b are the momenta of initial state particles. Finally, the action A is defined by

$$\begin{aligned}A[g] = &-\int d^4x\,d^4y K\psi(x)D^g_+(x-y)K\bar{\psi}(y) \\ &+ i\int d^4x[L(\psi) - L(\bar{\psi})],\end{aligned} \tag{1.3}$$

where $L(\psi)$ is the Lagrangian of the field theory in question (we shall restrict ourselves to the discussion of the field theory of a single scalar, Hermitian field). The Green's function

$$D^g_+(x) = (2\pi)^{-3}\int d^3p(2E)^{-1}\exp\{-ip\cdot x\}g(p) \tag{1.4}$$

can be substituted by $D^s_+(x)$ in the limit $g\to 1$. $D^s_+(x)$ is defined by (1.4), except the integral over the momentum is cut off at $|\mathbf{p}| = (s - 4m^2)^{1/2}$, where $s = (p_a + p_b)^2$. The substitution of $D_+(x) = D^1_+(x)$ by $D^s_+(x)$ is relevant only if we investigate the behavior of inclusive cross sections in the fragmentation region; for finite momenta the use of $D_+(x)$ is sufficient. Notice that the inclusive cross section is just the functional derivative of the generating functional $\sigma[g] = \sigma\langle 1\rangle_g$ with respect to $g(p)$.

The formalism described above can be put to use in various ways. Classical solutions of the field equations may dominate the expression of inclusive cross section (1.1). The classical solutions should satisfy the following set of integral equations [at $g(p) = g = \text{constant}$][1,2]

$$\begin{aligned}\psi(x) = &\int d^4yD_F(x-y)V(\psi(y)) \\ &+ ig\int d^4yD_+(x-y)V(\bar{\psi}(y)), \\ \bar{\psi}(x) = &\int d^4yD^*_F(x-y)V(\bar{\psi}(y)) \\ &- ig\int d^4yD^*_+(x-y)V(\psi(y)),\end{aligned} \tag{1.5}$$

where $V(\psi(x)) = \partial L_{\text{int}}(\psi)/\partial\psi(x)$, L_{int} is the interaction part of the Lagrangian and $D_F(x)$ is the casual Green's function of the massive Klein–Gordon equation, satisfying $KD_F(x) = -\delta^4(x)$. It is obvious that the function $\psi(x)$ and $\bar{\psi}(x)$ that satisfy Eq. (1.5) also satisfy the field equations

$$K\psi(x) = -V(\psi(x)), \quad K\bar{\psi}(x) = -V(\bar{\psi}(x)). \tag{1.6}$$

Note that no renormalization counterterms should be included in $V(\psi(x))$ in Eq. (1.5), because the semiclassical results are of lowest order in $\hbar$, while counterterms are not.

The contributions to inclusive cross sections discussed above are obviously nonperturbative. Contributions of perturbative origin can also be investigated in large orders (n) of the perturbation expansion with respect to the interacting part of action A of Eq. (1.3).[3,4] The leading contribution to the inclusive cross section is generated by instantonlike solutions of Eq. (1.5) with a changed coupling constant. Only the norm of these solutions depends on n in leading order of n^{-1}, consequently the n dependence of the Green's functions can be factored out and an estimate of the contribution of large order of perturbation theory to inclusive cross sections is obtained up to an unknown constant.[4]

[a]Permanent address: Department of physics, University of Cincinnati, Cincinnati, Ohio 45221, U.S.A.

 0022-2488/79/081720-07$01.00

The only specific case in which solutions of Eq. (1.5) have been studied in some detail up to now is the ψ^4 field theory,[1] in which the existence of $O(1,1)\times O(2)$ invariant solutions (depending on $p^2 = p_0^2 - \mathbf{p}^2$ and $p_1^2 = p_1^2 + p_2^2$ only) has been inferred from the investigation of Eq. (1.5) in momentum representation at large $p_\perp, p,\ p_\perp/p$ fixed. The existence of solutions of other symmetry groups could not be excluded, however, A solution that is $O(1,1)\times O(2)$ invariant leads to an inclusive cross section with Feynman scaling if we take into account the natural breaking of the symmetry for momenta of $O(s^{1/2})$, due to momentum conservation.

In the present paper we wish to prove that only $O(1,1)\otimes O(2)$ [or $O(1,1)$] invariant solutions of system (1.5) may exist. The breaking of the compact subgroup, $O(2)$, has no consequences on the symmetry properties of the inclusive cross section if no polarization is measured.

The proof is based on the detailed study of Eq. (1.5) in various regions of spacetime and various assumptions concerning the symmetry properties of the solution. We shall talk of $O(3,1)$ ($\psi,\bar\psi$ depends on x^2 only), $O(2,1)$ (dependence on x^2 and x_3), $O(1,1)\otimes O(2)$ (x^2 and $x_1^2 = x_1^2 + x_2^2$), $O(3)$ (x^2 and x_0) and $E(2)$ (x^2 and $x_+ = x_0 + x_3$) invariant solutions (ψ can always depend on sign (x_0) as well). The breaking of the compact part of the symmetry group does not influence the question of existence of solutions.

We shall define regions T^+ and T^- by $x^2 > 0$, $x_0 > 0$ and $x_0 < 0$, respectively. Region S is defined by $x^2 < 0$. Let us also define the Minkowski vectors $x_\parallel = (x_0, x_1, x_2)$ for $O(2,1)$ and $x_\parallel = (x_0, x_3)$ for $O(1,1)\otimes O(2)$. Then $x\in S^\pm$ if $x\in S$, $x_\parallel\cdot x_\parallel > 0$ and sign $(x_0) = \pm 1$, $x\in S^0$ if $x\in S$ and $x_\parallel\cdot x_\parallel < 0$. Finally we define $S^\pm$ for $E(2)$ by $x\in S$ and sign $(x_+) = \pm 1$.

We shall consider theories of a single Hermitian scalar field only. We shall assume that the interaction Lagrangian satisfies a few simple constraints. Defining $V(\psi) = dL_{\rm int}(\psi)/d\psi$, V is assumed to have the behavior of $V(\psi) = O(\psi^{N-1})$ for $\psi\to\infty$, where $N\leqslant 4$ (renormalizability constraint) and $V(\psi) = O(\psi^{n-1})$ for $\psi\to 0$, where $n\geqslant 3$ [the mass term is completely extracted from $L_{\rm int}(\psi)$]. The last constraint could be relaxed to $n > 2 + \epsilon$ spending considerably more effort. Finally, we shall assume that $W(\psi_1,\psi_2) = [V(\psi_1) - V(\psi_2)]/(\psi_1 - \psi_2)$ satisfies the conditions $W(\psi_1,\psi_2) = O(\psi_1^{N-2} + \psi_2^{N-2})$ for $\psi_1, \psi_2\to\infty$ and $W(\psi_1,\psi_2) = O(\psi_1^{n-2} + \psi_2^{n-2})$ for $\psi_1,\psi_2\to 0$. These and the previous constraints are satisfied by all of the models usually considered in the literature.

In the final part of the present section we shall cast Eq. (1.5) into a slightly more convenient form, as follows (we set $g = 1$ for the sake of simplicity):

$$\psi(x) = \int d^4y\, D_{\rm adv}(x-y)V(\psi(y)) + i\int d^4y D_+(x-y)[V(\bar\psi(y)) - V(\psi(y))], \tag{1.7}$$

$$\bar\psi(x) = \int d^4y\, D_{\rm adv}(x-y)V(\bar\psi(y)) + i\int d^4y D^*_+(x-y)[V(\bar\psi(y)) - V(\psi(y))]$$

We shall also make extensive use of the following equation [not independent of (1.7)]

$$\psi(x) - \bar\psi(x) = \int d^4y D_{\rm ret}(x-y)[V(\psi(y)) - V(\bar\psi(y))], \tag{1.8}$$

where $D_{\rm ret}(x)$ is defined as

$$D_{\rm ret}(x) = D_F(x) + iD^*_+(x) = \Theta(x_0)D(x^2), \tag{1.9}$$
$$D(z) = -(2\pi)^{-1}[\delta(z) - m\Theta(z)J_1(m\sqrt{z})/2\sqrt{z}].$$

Finally,

$$D_{\rm adv}(x) = D_{\rm ret}(-x).$$

The advantage of Eq. (1.7)–(1.8) in contrast to Eq. (1.5) is that their structure is made explicit, namely, it is obvious that they become Volterra integral equations in certain regions of spacetime and certain symmetries. More exactly $D_{\rm adv}(x-y)$ and $D_{\rm ret}(x-y)$ are Volterra kernels and the terms with $D_+(x-y)$ can be regarded as inhomogeneities. Another important and related property of Eq. (1.8) is that if $x\in T^-$, the $y\in T^-$ as well. Furthermore, if $x\in S$ then $y\in T^-$ or S. Equation (1.7) has a similar property with T^+ and T^- exchanged, provided $\bar\psi(x)\equiv\psi(x)$.

The proof of nonexistence of solutions is based on the simple observations made above. Every time $\psi(x)$ or $\psi(x) - \bar\psi(x)$ satisfies a *homogeneous* Volterra equation we can show after a study of the asymptotic behavior of the kernel that the equation has no nontrivial solutions. This circumstance is the consequence of the uniqueness of solutions of Volterra integral equations.[5] It is always sufficient to show that the solution is zero in a finite region around $x = 0$ or $1/x = 0$, since a solution of (1.6) may have singular points at $x = 0$ and ∞ only and it is identically zero if it vanishes on a finite subinterval of $(0,\infty)$.

One can put these considerations in a slightly different light by realizing that the admissible solutions of Eq. (1.6) are asymptotically dominated by the solutions of the free Klein–Gordon equation. A homogeneous Volterra equation, however relates the behavior of ψ on the left-hand side to the behavior of $V(\psi)$ on the right-hand side. Since $V(\psi)\sim O(\psi^{n-1})$ for $\psi\to 0$, the two sides do not match and the equation has no nontrivial solutions.

The organization of this paper is as follows. In Sec. II and III we prove the nonexistence of solutions for $O(3,1)$ symmetry and other symmetries [except of $O(1,1)\times O(2)$], respectively. In the conclusions (Sec. IV) we summarize our results, point out why $O(1,1)\otimes O(2)$ invariant solutions may exist and compare sectons with other known classical solutions, solitons and instantons. In the Appendix we investigate the asymptotic behavior of solutions of Eq. (1.6) near the singular points.

II. $O(3,1)$ SYMMETRIC SOLUTIONS

First we consider Eq. (1.7) for $O(3,1)$ invariant functions. The integrals over the angular variables can be performed exactly and one obtains the following equations for $\psi^+(x)$, $\psi^0(x)$ and $\psi^-(x)$ [the function $\psi(x)$ in T^+, S and T^-, respectively]:

$$\psi^+(x) = a_+ H_1^{(2)}(mx)/x - (\pi/2x)\int_x^\infty dyy^2 \times [J_1(mx)Y_1(my) - J_1(my)Y_1(mx)]V(\psi^+(y)) \tag{2.1}$$

$$\psi^0(x) = a_0 K_1(mx)/x - [I_1(mx)/x]\int_x^\infty dyy^2 K_1(my)V(\psi^0(y)) - [K_1(mx)/x]\int_0^x dyy^2 I_1(my)V(\psi^0(y)), \tag{2.2}$$

$$\psi^-(x) = a_- H_1^{(1)}(mx)/x + b_- J_1(mx)/x + c_- Y_1(mx)/x + (\pi/2x)\int_0^x dyy^2[J_1(mx)Y_1(my) - J_1(my)Y_1(mx)]V(\psi^-(y)) \tag{2.3}$$

$$a_+ = \int_0^\infty dyy^2\{(i\pi/4)H_1^{(1)}(my)[V(\bar\psi^+(y)) - V(\psi^+(y))] + (-i\pi/4)H_1^{(2)}(my)[V(\bar\psi^-(y)) - V(\psi^-(y))] + K_1(my)[V(\bar\psi^0(y)) - V(\psi^0(y))]\},$$

$$a_0 = (-2i/\pi)a_+ + \int_0^\infty dyy^2 J_1(my)V(\psi^+(y)), \tag{2.4}$$

$$a_- = -a_+,$$

$$b_- = (\pi/2)\int_0^\infty dyy^2 Y_1(my)V(\psi^+(y)) - 2\int_0^\infty dyy^2 K_1(my)V(\psi^0(y)),$$

$$c_- = (\pi/2)\int_0^\infty dyy^2 J_1(my)V(\psi^+(y)).$$

The equations for $\bar\psi(x)$ are obtained from Eqs. (2.1)–(2.4) by complex conjugation and the substitution $[\psi(x)]^* \to \bar\psi(x)$, $[\bar\psi(x)]^* \to \psi(x)$. We use the standard notations, $J_n(x)$, $Y_n(x)$, $H_n^{(1,2)}(x)$, $K_n(x)$ and $I_n(x)$ for Bessel functions.

First we prove that $a_0 = 0$. Suppose $a_0 \neq 0$, then $\psi^0(x) \sim x^{-2}$ for $x \to 0$. Indeed, none of the first two terms on the right hand side of Eq. (2.2) can compensate the term $a_0 K_1(mx)/x$; the second term is obviously smaller for $x \to 0$, while the first term would have the same behavior only if $V(\psi^0(x)) \sim x^{-4}$ for $x \to 0$. Such a behavior of $V(\psi^0(x))$ would make the second term divergent, so we proved that if $a_0 \neq 0$ then $\psi^0(x) \sim x^{-2}$ for $x \to 0$. On the other hand if $\psi^0(x) \sim x^{-2}$ then b_- of Eq. (2.4) would diverge. Consequently $a_0 = 0$.

The possible asymptotic behaviors of $\psi(x)$ at $x = 0$ are analyzed in the Appendix [Eqs. (A2) and (A4)]. They are $\psi(x) \sim x^{-2}$, $x^{-2/(N-2)}$ and x^0, where N is defined as $V(\psi) \sim g\psi^{N-1}$ for $\psi \to \infty$. The requirement of convergence of b_- excludes the behavior $x^{-2/(N-2)}$ as well. If $\psi(0)$ and $\bar\psi(0)$ are constants then a_i and b_i can be calculated, by making use of the field equations and integrating by parts in Eq. (2.4). We obtain

$$a_+ = [\bar\psi^+(0) + \bar\psi^-(0) - 2\bar\psi^0(0) - \psi^+(0) - \psi^-(0) + 2\psi^0(0)]/m,$$

$$a_0 = -(2i/\pi)a_+ = (2i/\pi)a_-, \qquad b_- = -2\psi^+(0)/m + 4\psi^0(0)/m, \qquad c_- = 0. \tag{2.5}$$

The vanishing of the coefficient $a_+ = i\pi a_0/2$ implies that $\psi^+(x) = 0$. The reason for the vanishing of $\psi^+(x)$ is that Eq. (2.1) becomes a homogeneous Volterra integral equation in variable x^{-1}. Suppose $n \geqslant 3$, where $V(\psi) \sim h\psi^{n-1}$ for $\psi \to 0$. then we obtain the following estimate for $\chi(x) = x^{3/2}\psi(x)$ {$\chi(x)$ is bounded, see Eqs. [(A2) and (A3)]}:

$$|\chi(x)| < (h\pi M^{n-2}/m)\int_x^\infty dyy^{-3(n-2)/2}|\chi(x)|, \tag{2.6}$$

where M is an upper bound for $|\chi(x)|$. Equation (4.6) was obtained by making use of the bounds $|J_1(z)| < z^{-1/2}$, $|Y_1(z)| < z^{-1/2}$, satisfied if $z > z_0$. The iteration of Eq. (2.6) shows that the only function satisfying this equation is $\chi(x) \equiv 0$.

The continuity of $\psi(x)$ as a function of x^2 at $x^2 = 0$ requires, however, that $\psi^+(0) = \psi^-(0) = \psi^0(0)$ (see Appendix). The asymptotic behavior $\psi^0(y) \to 0$ and $\psi^-(y) \to 0$ for $y \to 0$ is not admissible for an $O(3,1)$ invariant solution of Eq. (1.6), i.e., $\psi^-(x) \equiv \psi^0(x) \equiv 0$. In other words, no nontrivial global solution of Eqs. (2.1)–(2.4) exists.

III. $O(2,1)$, $E(2)$ and $O(3)$ INVARIANT SOLUTIONS

We shall follow a route, somewhat different from the one we have taken for the investigation of $O(3,1)$ invariant solutions to prove the nonexistence of solutions for $O(2,1)$, $E(2)$ and $O(3)$. We shall frequently rely on the Volterra character of the equations.

First we shall consider Eq. (1.8) for $x \in T^-$; having found that it is a Volterra equation for $\psi(x) - \bar\psi(x)$ we shall proceed to prove that $\psi(x) = \bar\psi(x)$ if $x \in T^-$. Then we move to region S^- and prove that $\psi(x) = \bar\psi(x)$ in this region as well. The investigation of S^- is followed by that of S^+ and T^+ proving in each region that $\psi(x) = \bar\psi(x)$. Using this result we move to Eq. (1.7) and investigate the same regions in reverse order. Each time we find the equation to be a Volterra equation for $\psi(x)$, and consequently we can prove that $\psi(x) = 0$. A slight complication arises in the $O(2,1)$ case, because the integral equation is not a Volterra equation if $x \in S^0$, but luckily the other regions decouple from S^0, i.e., if $x \in S^\pm$ or $T^\pm$ then $y \notin S^0$.

It is convenient to introduce the variables x_+ and x_- (similarly y_+ and y_-) as follows: $x_+ = x_0 + x_3$ for $E(2)$, $x_0 + |\mathbf{x}|$ for $O(3)$ and $x_3 + |\mathbf{x}|$ for $O(2,1)$, while $x_- = x \cdot x/x_+$ for $E(2)$, $x_0 - |\mathbf{x}|$ for $O(3)$ and $|\mathbf{x}| - x_3$ for $O(2,1)$. $\mathbf{x}$ is a three dimensional Euclidean vector for $O(3)$ and Minkowski vector for $O(2,1)$. Further on we introduce the notation

$$k(x_+, x_-; y_+, y_-) = \Theta(x_+ - y_+)\Theta(x_- - y_-) \times J_0(m[(x_+ - y_+)(x_- - y_-)]^{1/2}). \tag{3.1}$$

Using these notations the kernel of the equations,

$\int d^4y D_{\rm ret}(x-y)$ can be written as (after integrating over dummy variables)

$$\int d^4y D_{\rm ret}(x-y) \sim \int dy_+\, dy_-\, K(x_+,x_-;y_+,y_-), \tag{3.2}$$

where

$$K(x_+,x_-;y_+,y_-) \sim (y_+/x_+) k(x_+,x_-;y_+,y_-), \tag{3.3}$$

for $E(2)$

$$K(x_+,x_-;y_+,y_-) \sim [(y_+ - y_-)/(x_+ - x_-)]\Theta(y_+ - y_-) \times [k(x_+,x_-;y_+,y_-) - k(x_+,x_-;y_-,y_+)], \tag{3.4}$$

for $O(3)$,

$$K(x_+,x_-;y_+,y_-) = [(y_+ + y_-)/(x_+ + x_-)] k(x_+,x_-;y_+,y_-), \tag{3.5}$$

for $O(2,1)$.

Expressions for the kernel $\int d^4y D_{\rm adv}(x-y)$ are obtained by the substitution $x_+ \leftrightarrow y_+$, $x_- \leftrightarrow y_-$.

The similarity of the equations for the symmetry groups in consideration allows us to concentrate on one of these symmetries only. The proof of nonexistence of solutions can be readily extended to the other symmetry groups as well.

Let us take as the illustration of our method the case of $E(2)$ invariance. First we shall prove that $\psi(x) \equiv \bar\psi(x)$. Equation (1.8) for $x \in T^-$ reads as (we substitute $x_i \to -x_i$, $y_i \to -y_i$)

$$\Delta\psi(x_+,x_-) = \int_{x_+}^{\infty} dy_+ \int_{x_-}^{\infty} dy_- (y_+/x_+) \times J_0(m[(y_+ - x_+)(y_- - x_-)]^{1/2} \times W(\psi,\bar\psi)\Delta\psi(y_+,y_-), \tag{3.6}$$

where $\Delta\psi = \psi - \bar\psi$ and $W(\psi,\bar\psi) = [V(\psi) - V(\bar\psi)]/[\psi - \bar\psi]$.

We know that ψ, $\bar\psi$ and $\Delta\psi$ have the following asymptotic behavior for $x_+, x_- \to \infty$ [Eq. (A7)]

$$\Delta\psi \sim C_1(r) \sin mx/x^{3/2} + C_2(r) \cos mx/x^{3/2}, \tag{3.7}$$

where $r = x_+/x_-$ and $C_i(r)$ are polynomially bounded. Any function $\Delta\psi$ that satisfies (1.6) and decreases faster than the expression given in (3.7) must be identically zero.

Since $W(\psi,\bar\psi) = (\psi^{n-2} + \bar\psi^{n-2}) = O(\psi + \bar\psi)$ a typical term on the right hand side of Eq. (3.6) gives the following contribution at large values of $x_+ = x_- = x$ [we set $y_+ = x(1+\eta_+^2)$, $y_- = x(1+\eta_-^2)$]

$$\Delta\psi(x,x) \sim x^{-1} \int_0^{\infty} d\eta_+ \eta_+ \int_0^{\infty} d\eta_- \eta_- (1+\eta_+^2)^{-1/2} \times (1+\eta_-^2)^{-3/2} P(mx[(1+\eta_+^2)(1+\eta_-^2)]^{1/2}) \times C((1+\eta_+^2)/(1+\eta_-^2)), \tag{3.8}$$

where P is a periodic function (maybe a constant) of its argument and $C(z)$ is such a function that the integrals over η_+ and η_- are finite.

Only small values of η_+ and η_- contribute to the integral on the right hand side of Eq. (3.8) if P is definitely not a constant. We obtain [say $P(z) = \sin z$]

$$\begin{aligned}\Delta\psi(x,x) &\sim x^{-1} \int_0^{\infty} d\eta_+ \eta_+ \int_0^{\infty} d\eta_- \eta_- J_0(mx\eta_+\eta_-) \times \sin(mx + mx\eta_+^2/2 + mx\eta_-^2/2) C(1) \\ &= x^{-3} \int_0^{\infty} d\eta_+ \eta_+ \int_0^{\infty} d\eta_- \eta_- J_0(m\eta_+\eta_-) \times \sin(mx + m\eta_+^2/2 + m\eta_-^2/2) C(1) \\ &= a \sin(mx)/x^3 + b\cos(mx)/x^3,\end{aligned}$$

contradicting our previous remark concerning the behavior of $\Delta\psi$ for large x.

If P is a constant then we use the standard trick of writing

$$J_0(z\eta_+) = \int_0^{\infty} d\varphi \sin(z\eta_+ \cosh\varphi),$$

then we extend the integral over η_+ to $-\infty < \eta_+ < \infty$ and deform the contour of integration to pick up the singularity of the integration over η_+ at $\eta_+ = -i$.

One obtains the following expression for the right hand side of Eq. (3.8)

$$\begin{aligned}\Delta\psi(x,x) &\sim x^{-1} \int_1^{\infty} d\rho\,\rho \int_0^{\infty} d\eta_- \eta_- \times f(\rho, 1+\eta_-^2) K_0(mx\eta_-\rho) \\ &\sim x^{-3} \int_1^{\infty} d\rho\,\rho \int_0^{\infty} d\eta\eta f(\rho,1) K_0(m\eta\rho),\end{aligned} \tag{3.9}$$

where we substituted $\rho = i\eta_+$. If the singularity of $C((1+\eta_+^2)/(1+\eta_-^2))(1+\eta_+^2)^{-1/2}$ at $\eta_+ = -i$ is such that the integral on the right hand side of Eq. (3.9) does not converge at $\rho = 1$, then the choice of the contour of the integral over variable η_+ is slightly different: It does not go through the point $\eta_+ = -i$, but still at every point $\mathrm{Im}\eta_+ > \epsilon > 0$. The behavior of $\psi(x_+,x_-)$ at large x would still be $\psi(x) \sim x^{-3}$. Consequently, $\psi(x) = \bar\psi(x)$ if $x \in T^-$.

Using the result $\psi(x) = \bar\psi(x)$ for $x \in T^-$ we can write down integral equation (1.8) for $x \in S^-$, and we get automatically $y \in S^-$. Introducing variable $x = (x_+x_-)^{1/2}$, $s = x_+/x_-$, $(y_+y_-)^{1/2} = ux$, $y_+/y_- = \rho s$ we obtain

$$\Delta\psi(x,s) \sim x^2 \left[\int_1^{\infty} du u^2 \int_u^{\infty} d\rho + \int_0^1 du u^2 \int_{1/u}^{\infty} d\rho \right] \cdot J_0(m[u(\rho + \rho^{-1}) - 1 - u^2]^{1/2}) \times W(\psi,\bar\psi)\Delta\psi(xu, s\rho). \tag{3.10}$$

Since s appears only in the argument of functions ψ and $\bar\psi$ on the right hand side of Eq. (3.10) we immediately realize that a consistent solution of this equation is possible only if $\bar\psi \sim \psi \sim$ const for $s \to \infty$, while $\psi - \bar\psi = \Delta\psi \sim S^{-\alpha}$, where the value of α should be calculated from the substitution of the trial function, $\Delta\psi = \eta(x) s^{-\rho}$ into Eq. (3.10). On the other hand, if $\psi(x,s) \to \psi(x)$ (independent of s) for large s, then the integral equation reduces to the $O(3,1)$ invariant equation, and we know that no $O(3,1)$ invariant solutions exist because the condition of continuity can not be satisfied at $x^2 = 0$.

The proof of nonexistence of solutions in S^+ is identical with the proof for $x \in S^-$ provided x^+ and y^+ are exchanged with x^- and y^-.

Finally, if $\Delta\psi = 0$ for $x \in T^-, S$ then Eq. (1.8) reads for $x \in T^+$ as follows:

$$\Delta\psi(x_+, x_-) = \int_0^{x_+} dy_+ \int_0^{x_-} dy_- (y_+/x_+) \times kW(\psi,\bar{\psi})\Delta\psi(y_+, y_-). \tag{3.11}$$

It is possible to set $k = 1$ for small enough x_+ and x_-. The asymptotic behavior of ψ for small x_+ and x_- is either x_+^{-1} x_-^m or x_+^m or $(x_+ x_-)^{-1/(N-1)}$; in either case the x_+ and x_- dependences factorize for sufficiently small x_- [See Eqs. (A4) and (A8)]. A similar statement is valid for $\Delta\psi$ as well. Consequently, we have the equation

$$\Delta\psi(x_-) = C \int_0^{x_-} dy_- W(\psi,\bar{\psi})\Delta\psi(y_-) \tag{3.12}$$

that implies $\Delta\psi(x_-) \to 0$ for $x_- \to 0$. On the other hand $\Delta\psi(x)$ satisfies a differential equation $\Delta\psi = CW\Delta\psi$ the solution of which is finite and nonzero at x_- if $W(\psi,\bar{\psi})$ is bounded and infinite if $W(\psi,\bar{\psi})$ is unbounded, implying that Eqs. (3.12) and (3.11) have only trivial solutions.

Having proven $\psi(x) = \bar{\psi}(x)$ one may turn to Eq. (1.7) and consider regions T^+, S^+, S^-, and T^-, each after the other. Thus, one obtains equations analogous to (3.6), (3.10), and (3.11) except of the substitutions $\Delta\psi \to \psi$, $W(\psi,\bar{\psi})\Delta\psi \to V(\psi)$. The absence of nontrivial solutions can be shown by methods quite analogous to the ones used to analyze the solutions of the original equations.

The methods we applied to prove the nonexistence of solutions of Eq. (1.8) and (1.7) in regions S^+, S^-, T^+, and T^- would not apply to region S^0 for $O(2,1)$ invariant solutions. Let us take, however, $x \in S^+$ and $y \in S^0$, then we obtain (we choose $x_0 = |\mathbf{x}|$, $|\mathbf{x}| \equiv \mathbf{x}_\parallel$)

$$\int D_{\text{ret}}(x-y)d^4y \sim \int_0^\infty dy_\parallel (y_\parallel/x_\parallel) dy_3 \int_0^\infty du D(u) \tag{3.13}$$

where $u = -2x_\parallel y_\parallel \sinh\alpha + x_\parallel^2 - y_\parallel^2 - (x_3 - y_3)^2$ and $D(z)$ was defined by Eq. (1.9). It is easy to show that

$$\int_0^\infty du D(u) = 0,$$

consequently the regions S^+ (similarly T^+) and S^0 decouple. Thus, we can prove that $O(2,1)$ invariant solution $\psi(x)$ has to disappear for $x \in T^+, T^-, S^+, S^-$. The Fourier transform of a function $\psi(x)$ that differs from zero only if $x \in S^0$ cannot, however, have a pole at $p^2 = m^2$ unless it increases exponentially for $x \to \infty$. Such a behavior is clearly not admissible by the integral equation for $x \in S^0$. In other words, Eq. (1.5) has no $O(2,1)$ invariant secton solutions, solutions that contribute to inclusive cross sections. One can also remark that a solution differing from zero only if $x \in S^0$ does not satisfy the condition of continuity at $x_\parallel^2 \equiv x_0^2 - x_1^2 - x_2^0 = 0$, as discussed in the Appendix and as such it is not a global solution.

Finally, we shall discuss solutions having singularities at locations other than $\lambda = 0$. Such solutions, (or rather $\Delta\psi$) also disappear at large x^2, for $x \in T^-$. A solution of (1.6), which is zero for $x > c$, $c \neq 0$ cannot be different from zero for $x < c$, however, since its behavior at $x \lesssim c$ could only be $\psi \sim c - x$ or $\psi \sim$ constant and such a behavior does not provide a global solution of Eq. (1.6) (see the last part of the Appendix). Consequently, the solution is zero everywhere. The proof outlined above loses its validity if the singularity, at a point other than $x = 0$, is a consequence of a translation $x_\mu \to x_\mu + c_\mu$, but this case trivially reduces to the problem in which there are singularities at $x = 0$ only, after the inverse translation.

IV. CONCLUSIONS

The nonexistence of solutions of integral equation (1.5) was proven in the previous two sections except for possible solutions of $O(1,1) \times O(2)$ [or $O(1,1)$] symmetry. The proof was based on the recognition of the simple fact that the integral equation reduces to a Volterra equation in all regions and symmetry groups, except of region S or S^0 for $O(3,1)$ or $O(2,1)$ and $O(1,1) \times O(2)$ symmetries. The part of the proof concerning regions S^+ and S^- relies on the fact that no $O(3,1)$ invariant solutions exist.

$O(2,1)$ invariant solutions were eliminated by noticing the decoupling of region S^0 from other regions in the kernel of the integral equation.

Finally, we proved that $O(3,1)$ invariant solutions do not exist for two reasons:

(1) The coefficients of the inhomogeneous terms in Eq. (2.1)–(2.3) a_0, a_+ and a_- have to disappear, otherwise $\psi^0(x)$ is too singular for the convergence of some integrals (this circumstance excludes the solution, singular at $x = 0$, $\psi(x) \sim x^{-2}$);

(ii) The continuity of the solution at $x^2 = 0$ eliminates the regular solution [$\psi(x) \sim$ const for $x \to 0$] as well.

We shall point out below that our arguments used to eliminate solutions of other symmetries are not applicable to the case of $O(1,1) \otimes O(2)$ symmetry.

One can follow the methods applied in Sec. III to prove the relation $\psi(x) = \bar{\psi}(x)$ if $x \in T^-, S^-$ for $O(1,1) \otimes O(2)$ symmetry as well. On the other hand, the kernel of the equation is not Volterra if $x, y \in S^0$, since we have

$$\int d^4y D_{\text{ret}}(x-y) = \int dy_\parallel^2 \int dy_\perp^2 K(x_\parallel, x_\perp; y_\parallel, y_\perp) \sim \int dy_\parallel^2 \int dy_\perp^2 \int_0^\infty d\alpha \times \int_{-\pi}^{\pi} d\vartheta D(-x_\parallel^2 - y_\parallel^2 + 2x_\parallel y_\parallel \cosh\alpha - x_\perp^2 - y_\perp^2 + 2x_\perp y_\perp \cos\vartheta) \tag{4.1}$$

and no matter how we choose $x_\parallel, x_\perp, y_\parallel$ and $y_\perp$ we can find α and ϑ such that argument of $D(z)$ [see Eq. (1.9)], $z > 0$ and we have a nonzero contribution. Unfortunately, we could not express kernel K in terms of known special functions (except for the case $m = 0$, when K can be expressed by elliptic functions).

The argument used to eliminate $O(2,1)$ invariant solutions does not apply here, since the kernel of the equation for $x \in S^+, y \in S^0$ has the form

$$\int d^4y D_{\text{ret}}(x-y)$$

$$= \int dy_{\|}^2 \int dy_{\perp}^2 K(x_{\|}, x_{\perp}; y_{\|}, y_{\perp})$$

$$\sim \int dy_{\|}^2 \int dy_{\perp}^2 \int_{-\infty}^{\infty} d\alpha \int_{-\pi}^{\pi} d\vartheta D(x_{\|}^2 - y_{\|}^2 - 2x_{\|} y_{\|} \sinh\alpha - y_{\perp}^2 - x_{\perp}^2 + 2x_{\perp} y_{\perp} \cos\vartheta), \quad (4.2)$$

and the integrals over α and ϑ give a nonzero result.

Further on we can show that the arguments used to eliminate $O(3,1)$ invariant solutions cannot be used to prove the nonexistence of $O(1,1) \otimes O(2)$ invariant solutions.

(i): The inhomogeneity, analogous to the term $a_0 K_1(mx)/x$ in Eq. (2.2) is much less singular; it behaves like $\log x_{\|}$ for $x_{\perp}, x_{\|} \to 0$. Consequently, even if the inhomogeneity is nonzero, $\psi(x)$ is not too singular to make all of the appearing integrals convergent;

(ii) An $O(1,1) \otimes O(2)$ invariant solution of Eq. (1.6) can have a finite jump at $x_{\|}^2 = 0$, so $\Delta\psi = 0$ for $x \in T^-, S^-$ and $\Delta\psi \neq 0$, $\Delta\psi \to$ const for $x_{\|}^2 \to 0$, $x \in S^0$ is an admissible global solution of Eq. (1.6). This phenomenon is restricted to solutions with a symmetry group, the noncompact part of which is $O(1,1)$ (see Appendix).

The question of existence of $O(1,1) \otimes O(2)$ invariant solutions of Eq. (1.5) will be investigated in future publications.

Finally, we shall make some remarks concerning the relation of secton solutions on one hand and soliton or instanton solutions on the other. The most crucial difference between secton and soliton solutions lies in the fact that only secton solutions possess mass shell singularities (except possibly of the mass = 0 case). The Fourier transform of a soliton solution contains a factor $\delta(p_0)$, so that $p_0^2 = \mathbf{p}^2 + m^2$ cannot be satisfied. Instanton solutions satisfy Eq. (1.5) in region S (or S^0) provided $\psi(x) = \bar{\psi}(x)$, but they do not provide a global solution of the integral equation. As we learned, they cannot even be extended to global solutions except maybe in the case $O(1,1) \otimes O(2)$. The reason of the difference of secton and instanton solutions is that the latter satisfy an integral equation obtained from (1.5) if the terms with $D_+(x-y)$ functions are omitted.

Thus, one obtains an integral equation the kernel of which if an *analytic* function of x_0 and one can continue to Euclidean metric. The D_+ functions are not analytic, however, and when solving Eq. (1.5) one has to stay in Minkowski space, find and match solutions for time and spacelike coordinates as well.

ACKNOWLEDGMENTS

The author has benefitted from stimulating discussions with Dr. I.G. Halliday and Dr. J.J. Scanio, and is grateful to Professor T.W.B. Kibble for the hospitality at the Blackett Laboratory, where a major part of this work was completed.

The author also wishes to thank the Science Research Council (G.B.) and the U.S. Department of Energy under contract number EY-76-S-02-2978 for partial support.

APPENDIX

In this Appendix we shall investigate the asymptotic behavior of solutions of Eq. (1.6) near the singular points.

A. $O(3,1)$ invariant solutions

Equation (1.6) reads for $O(3,1)$ invariant solutions as follows ($x^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$, $x^2 > 0$)

$$[d^2/dx^2 + (3/x)d/dx + m^2]\psi(x) = -V(\psi(x)). \quad (A1)$$

The possible asymptotic behavior of the solutions of Eq. (A1) for $x \to \infty$ depend on the behavior of $V(\psi)$ for $\psi \to 0$. We shall assume that $V(\psi) \sim g\psi^{n-1}$ for $\psi \to 0$, where $n > 2 + \epsilon$, $\epsilon > 0$. Then the general solution of Eq. (A1) is dominated by the general solution of the corresponding linear equation for $x \to \infty$, namely

$$\psi(x) \sim c_1 Y_1(mx)/x + c_2 J_1(mx)/x. \quad (A2)$$

The special solution $\psi(x) \to C$ for $x \to \infty$, where $m^2 C + V(c) = 0$ does not satisfy Eq. (2.1) and leads to infinite action. $\psi(x)$, as given by Eq. (A2) satisfies the constraint

$$|x^{3/2}\psi(x)| < N \quad \text{if } x \geqslant \epsilon > 0. \quad (A3)$$

The behavior of $\psi(x)$ near $x = 0$ is either dominated by the solution of the linear part of Eq. (A1), in which case $\psi(x) \sim d_1 x^{-2} + d_2$ for $x \to 0$, or one may have a special solution that satisfies the nonlinear equation only

$$\psi(x) \sim x^{-2/(N-2)}. \quad (A4)$$

The $\psi(x) \sim x^{-2}$ behavior is possible only if $N < 3$.

Finally, we shall examine the behavior of $\psi(x)$ near $x^2 \equiv \lambda = 0$, provided ψ is bounded. We can introduce variable λ instead of x in Eq. (A1) to obtain

$$(4\lambda d^2/d\lambda^2 + 8d/d\lambda + m^2)\psi(\lambda) = -V(\psi(\lambda)). \quad (A5)$$

$\psi(\lambda)$ has to be continuous at $\lambda = 0$. Suppose $\psi(\lambda)$ contains a term proportional to $\theta(\lambda) f(\lambda)$, then one obtains the following term on the left hand side of Eq. (A5)

$$4\lambda\delta'(\lambda) f(\lambda) + 8\delta(\lambda) f(\lambda) = 4\delta(\lambda) f(0),$$

and consequently we do not have a global solution of Eq. (A5).

B. $E(2)$ invariant solutions

Let us introduce the variables $x_+ = x_0 + x_3$, $x_- = x^2/x_+$ and $r^2 = x_-/x_+ = x^2/x_+^2$. Equation (1.6) has the following form for an $E(2)$ invariant solution in terms of the variables introduced above:

$$\{\partial^2/\partial x^2 + (3/x)\partial/\partial x - [r^2\partial^2/\partial r^2 + 3r\partial/\partial r]/x^2 + m^2\} \times \psi(x,r) = -V(\psi(x,r)) \quad (A6)$$

It is obvious that at fixed x, $r \to \infty$, $N > 2$, $\psi(x,r) \to 0$ or $\psi(x,r) \to \chi(x)$, independent of r, otherwise the right-hand side of Eq. (A6) could not be compensated by the left-hand side. On the other-hand solution of the linear part of the equation has to dominate the asymptotic behavior of the solution for $x \to \infty$, just like in the $O(3,1)$ invariant case. The general solution of the linear part of the equation is

$$\psi(x,r) \sim Y_t(mx)x^{-1}(c_1 r^{-1-t} + c_2 r^{-1+t}) + J_t(mx)x^{-1}(d_1 r^{-1-t} + d_2 r^{-1+t}) \sim C_1(r)x^{-3/2}\sin(mx) + C_2(r)x^{-3/2}\cos(mx), \quad (A7)$$

with t arbitrary.

The behavior of $\psi(x,r)$ for $x \to 0$ is either controlled by

the solution of the linear part of Eq. (1.6) in which case [see Eq. (A7)]

$$\psi(x,r)\sim c_1(xr)^t + c_2(xr)^{-1}(x/r)^t = c_1 x_+^t + c_2 x_+^{-1} x_-^t \tag{A8}$$

(the condition of dominance of the linear part sets some obvious constraints on Ret), or it is controlled by the interplay of the nonlinear and linear parts, in which case the solution becomes $O(3,1)$ invariant for $x\to 0$ and its behavior is given by Eq. (A4).

C. Other symmetries

Equation (1.6) takes the following form for an $O(k,1)\otimes O(3-k)$ invariant solution $(k=0,1,2)$

$$\{\partial^2/\partial x_\parallel^2 + (k/x_\parallel)\partial/\partial x_\parallel - \partial^2/\partial x_\perp^2 - [(2-k)/x_\perp]\partial/\partial x_\perp + m^2\}\psi(x_\parallel,x_\perp) = -V(\psi(x_\parallel,x_\perp)), \tag{A9}$$

where $x_\parallel$ is the norm of the $O(k,1)$ vector, and $x_\perp$ is that of the $O(3-k)$ vector.

We shall not discuss the asymptotic behavior of $\psi(x_\parallel,x_\perp)$ at the singular points in details, the discussion and results are substantially similar to those in the $E(2)$ invariant case.

The only question we wish to examine is the behavior of the solutions near $\lambda \equiv x_\parallel^2 = 0$. Equation (A9) in terms of variables λ and $x_\parallel$ reads as follows:

$$\{4\lambda\partial^2/\partial\lambda^2 + (2k+2)\partial/\partial\lambda - \partial^2/\partial x_\perp^2 - [(2-k)/x_\perp]\partial/\partial x_\perp + m^2\}\psi(\lambda,x_\perp) = -V(\psi(\lambda,x_\perp)). \tag{A10}$$

Suppose that $\psi(\lambda,x_\perp)$ has a finite jump at $\lambda=0$, i.e., $\psi(\lambda,x_\perp)$ contains a term $\theta(\lambda)f(\lambda,x_\perp)$, with $f(\lambda,x_\perp)$ finite and nonzero at $\lambda=0$. We obtain the following contribution on the left-hand side of Eq. (A10)

$$\begin{aligned}&[4\lambda\partial^2/\partial\lambda^2 + (2k+2)\partial/\partial\lambda]\theta(\lambda)f(\lambda,x_\perp)\\ &\quad\sim 4\lambda\delta'(\lambda)f(\lambda,x_\perp) + 8\lambda\delta(\lambda)f'(\lambda,x_\perp) + (2k+2)\delta(\lambda)\\ &\qquad\times f(\lambda,x_\perp)\\ &\quad\sim(2k-2)\delta(\lambda)f(0,x_\perp).\end{aligned} \tag{A11}$$

It is obvious that the finite jump is admissible only for $k=1$, or with other words for an $O(1,1)\otimes O(2)$ invariant solution.

D. Singularities at points $x^2\neq 0$

The field equation for $E(2)$ invariant solutions can be written as follows:

$$[4\lambda\partial^2/\partial\lambda^2 + 8\partial/\partial\lambda + 4x_+\partial^2/\partial x_+\partial\lambda]\psi(x_+,\lambda) = -V(\psi(x_+,\lambda)), \tag{A12}$$

where $\lambda = x^2$, $x_+ = x_0 + x_3$. Suppose $\psi = 0$ for $\lambda > c$. Then the behavior of $\psi(\lambda,x_+)$ near $\lambda = c$ is governed by the term $(4\lambda\partial^2/\partial\lambda^2)\psi$ of the equation, unless $c = c'x_+$, in which case the term $(4x_+\partial^2/\partial x_+\partial\lambda)\psi$ is equally important. Disregarding the last case, in which a linear transformation translates the singularity to $\lambda = 0$, with an unchanged field equation, the behavior of $\psi(\lambda,x_+)$ near $\lambda = c$ could only be $\psi\sim\theta(c-x)$ or $\psi\sim\theta(c-x)(c-x)$, generating terms of the type $\delta'(c-x)$ and $\delta(c-x)$, respectively, on the left hand side of Eq. (A12).

[1] P. Suranyi, Phys. Rev. D **17**, 1160 (1978).
[2] P. Suranyi, Phys. Rev. D **18**, 2130 (1978).
[3] L.N. Lipatov, JETP **72**, 411 (1977).
[4] E. Brezin, J.C. Le Guillou and J. Zinn-Justin, Phys. Rev. D **15**, 544 (1977).
[5] R. Miller, *Nonlinear Volterra Integral Equations* (Benjamin, Menlo Park, California, 1971).

Scattering of a beam of particles by a potential

John D. Dollard and Charles N. Friedman
Math Department, The University of Texas, Austin, Texas 78712

(Received 22 September 1978; revised manuscript received 11 November 1978)

We discuss the potential scattering into a cone C in position space of a beam of particles with localized momenta distributed over a set K, and the derivation from time-dependent scattering theory of the formula
$\int_K d\tilde{k} \int_C d\Omega\, \sigma_{\tilde{k}}(\Omega)$
[where $\sigma_{\tilde{k}}(\Omega)$ is the scattering cross section] as a measure of the probability of this scattering. We make use in our discussion of a result of Dollard giving the probability of scattering into a cone for a single wavepacket.

1. INTRODUCTION

In the time-independent approach to scattering theory one begins with the time-independent Schrödinger equation

$$(\Delta + k^2)\psi(\tilde{x}) = V(\tilde{x})\psi(\tilde{x}), \tag{1.1}$$

where $V(\tilde{x})$ is, say, a potential which tends to 0 as $|\tilde{x}| \to \infty$, and one looks for solutions having the asymptotic form

$$\psi_{\tilde{k}}(\tilde{x}) \sim e^{i\tilde{k}\cdot\tilde{x}} + f_{\tilde{k}}(\Omega)\frac{e^{ikr}}{r}, \tag{1.2}$$

as $r = |\tilde{x}| \to \infty$. The solution $\psi_{\tilde{k}}(\tilde{x})$ is supposed to be related to a scattering process in which an incoming plane wave, $e^{i\tilde{k}\cdot\tilde{x}}$, interacts with the potential, resulting in another plane wave with the same momentum (corresponding to part of the original plane wave which is not scattered) and an outgoing cloud of scattered particles [represented by the second term in (1.2)]. The quantity $f_{\tilde{k}}(\Omega)$ is called the scattering amplitude, and the quantity,

$$\sigma_{\tilde{k}}(\Omega)d\Omega = |f_{\tilde{k}}(\Omega)|^2 d\Omega, \tag{1.3}$$

is the (differential) scattering cross section and is supposed to describe the number of particles scattered per unit time into the solid angle $d\Omega$, divided by the flux of the incident particles. Various interesting quantities are then expressible in terms of the cross section $\sigma_{\tilde{k}}(\Omega)$. For example the authors of Ref. 1, using both rigorous mathematical arguments of time-dependent scattering theory and some plausibility arguments, arrive at essentially the following result. Suppose we consider a "beam of particles" with momentum uniformly distributed in some subset K of momentum space which is scattered by the (radially symmetric) potential V. Then the probability of scattering into a cone C in position space is proportional to

$$\int_K d\tilde{k} \int_C d\Omega \sigma_{\tilde{k}}(\Omega), \tag{1.4}$$

at least if one "ignores the forward scattering." A similar result appears in Refs. 2 and 3.

Of course one would like to rigorously derive such results using the apparatus of time-dependent scattering theory; in carrying out such a program one would deal with *normalizable* wavefunctions and study their time evolution as they interact with the potential. This is not to say that plane waves or solutions of the Schrödinger equation of the form (1.2) should not be utilized; these are useful mathematical objects which appear in the discussion of the questions we have mentioned above, and it is only their *interpretation* as "particles" or "beams of particles" that should be eliminated.

In the present article we will derive a result related to (1.4), from the time-dependent theory. Our starting point will be the following result of Dollard[4]: given an initial state u of a particle, the probability $P(u,c)$ that at large positive times the particle will be found in the cone C (i.e., the probability that the particle scatters into C) is given by

$$P(u,c) = \int_C d\tilde{k}' \, |\widehat{\mathrm{Su}}(\tilde{k}')|^2 \tag{1.5}$$

(where appropriate assumptions on the potential V are made to insure that the S operator exists, and $\widehat{\mathrm{Su}}$ is the Fourier transform of Su). One might then try to recover the formula (1.4) by taking for u in (1.5),

$$u(\tilde{x}) = \mathrm{const} \int_K e^{i\tilde{k}\cdot\tilde{x}} d\tilde{k}, \tag{1.6}$$

and expressing S in terms of the scattering amplitude $f_{\tilde{k}}(\Omega)$. However, this does not lead to (1.4); the problem is that although (1.6) does represent a particle with momentum uniformly distributed over K, it is not the only such wavefunction with such an interpretation. One could just as well take

$$u(\tilde{x}) = \mathrm{const} \int_K e^{i\tilde{k}\cdot\tilde{x}} e^{i\phi(\tilde{k})} d\tilde{k}, \tag{1.7}$$

and to obtain anything like (1.4) it seems necessary to take the u in (1.5) to have the form (1.7) and then average over the phase $\phi(\tilde{k})$ in some way. Since the possible phase functions form an infinite dimensional space, the nature of such an averaging process is somewhat problematical. The usual solution to this problem (e.g., Rev. 5, has been to consider only those $\phi(\tilde{k})$ of the form

$$\phi(\tilde{k}) = \tilde{\xi}\cdot\tilde{k}, \quad \tilde{\xi}\cdot\tilde{k}_0 = 0,$$

where $\tilde{\xi}$ is a vector in $\mathbb{R}^3$ and $\tilde{k}_0$ is some fixed vector in K. Then the average over phases amounts to a two-dimensional integration over the plane perpendicular to $\tilde{k}_0$. As might be expected, this leads to an answer which depends in general

 0022-2488/79/081727-04$01.00

on $\tilde{k}_0$. The dependence on $\tilde{k}_0$ disappears if we let K shrink down to the point $\tilde{k}_0$, but then the original idea of a beam of particles with momentum distributed over K also disappears. The main problem here seems to be that wavefunctions such as (1.7) represent a single particle, not a beam of particles. We will overcome this difficulty by actually considering a many-particle wavefunction in our calculation. We will also take into account the possibility of forward scattering and will derive a result which reduces to (1.4) in the region where forward scattering does not occur (i.e., when K and C are disjoint).

2. SOME TECHNICAL PRELIMINARIES

In this section we will discuss several results concerning the existence of the scattering operator and solutions of the Schrödinger equation of type (1.2), and the relationship between these objects, which holds for certain classes of potentials $V(\tilde{x})$ defined on three-dimensional space $\mathbb{R}^3$. For the sake of simplicity we will not describe the most general potentials for which these results are known; we refer to Ref. 6 for further results.

In order to use the result of Dollard mentioned in the Introduction we shall need existence of the scattering operator. It is known[7] that for $V(\tilde{x}) \in L^2(\mathbb{R}^3) \cap L^1(\mathbb{R}^3)$ the Moller wave operators,

$$\Omega^{\pm} = \underset{t \to \pm\infty}{\text{s-lim}}\ e^{itH} e^{-itH_0}, \tag{2.1}$$

exists on $L^2(\mathbb{R}^3)$. Here

$$H_0 = -\Delta = -\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}\right), \tag{2.2}$$

$$H = H_0 + V.$$

Furthermore, the S operator,

$$S = (\Omega^+)^* \Omega^-, \tag{2.3}$$

is unitary.

We next discuss solutions of the integral equation,

$$\psi_{\tilde{k}}(\tilde{x}) = e^{i\tilde{k}\cdot\tilde{x}} - \frac{1}{4\pi}\int d\tilde{y}\, \frac{e^{ik|\tilde{x}-\tilde{y}|}}{|\tilde{x}-\tilde{y}|} V(\tilde{y})\psi_{\tilde{k}}(\tilde{y}); \quad k = |\tilde{k}|, \tag{2.4}$$

which is formally equivalent to (1.1) together with the requirement that the solution behave asymptotically like a plane wave plus an outgoing spherical wave. Ikebe[8,9] has proved that if $V(\tilde{x})$ is real, locally Hölder continuous except at a finite number of singularities,

$$V \in L^2(\mathbb{R}^3) \text{ and } V(\tilde{x}) = O(|\tilde{x}|^{-3-\epsilon})\ \epsilon > 0 \text{ as } |\tilde{x}| \to \infty, \tag{2.5}$$

then for $\tilde{k} \neq \tilde{O}$, (2.4) has a solution $\psi_{\tilde{k}}(\tilde{x})$ which is bounded and uniformly continuous in $\tilde{x}$ and $\tilde{k}$ for $\tilde{k}$ ranging over any compact set excluding $\tilde{O}$. Furthermore, $\psi_{\tilde{k}}(\tilde{x})$ has the asymptotic form

$$\psi_{\tilde{k}}(\tilde{x}) = e^{i\tilde{k}\cdot\tilde{x}} - \frac{1}{4\pi}\frac{e^{ik|\tilde{x}|}}{|\tilde{x}|}\int d\tilde{y}\, e^{-i\tilde{k}'\cdot\tilde{y}} V(\tilde{y})\psi_{\tilde{k}}(\tilde{y}) + o\left(\frac{1}{|\tilde{x}|}\right), \tag{2.6}$$

$$\tilde{k}' = k(\tilde{x}/|\tilde{x}|),$$

which is of the form (1.2) if we put

$$f_{\tilde{k}}(\Omega) = -\frac{1}{4\pi}\int d\tilde{y}\, e^{-i\tilde{k}'\cdot\tilde{y}} V(\tilde{y})\psi_{\tilde{k}}(\tilde{y}). \tag{2.7}$$

In (2.7) Ω denotes the polar angles of $\tilde{k}'$ with $\tilde{k}$ taken as the polar axis ("z axis"). We will also need to know the relation between the S operator and $f_{\tilde{k}}(\Omega)$ of (2.7). Ikebe[9] proves that for $u \in L^2(\mathbb{R}^3)$ we have

$$\widehat{Su}(\tilde{k}') = \hat{u}(\tilde{k}') + \frac{1}{\pi}\int d\tilde{k}\, \hat{u}(\tilde{k}) f_{\tilde{k}}(\Omega)\delta(k^2 - k'^2), \tag{2.8}$$

where $\hat{}$ denotes the Fourier transform and δ is the one-dimensional delta distribution.

Remark: Our notation differs slightly from that of Ikebe, e.g., in the choice of constants in the definition of $f_{\tilde{k}}(\Omega)$. Also, Ikebe's form of (2.8) does not contain the $\delta(k^2 - k'^2)$ which may be removed by an integration.

We now use (2.8) to express the probability $P(u,c)$ in (1.5) in terms of the scattering amplitude $f_{\tilde{k}}(\Omega)$ for a particular u of the form

$$u(\tilde{x}) = (2\pi)^{-3/2}|K|^{-1/2}\int_K d\tilde{k}\, e^{i\tilde{k}\cdot\tilde{x}}, \tag{2.9}$$

where $K \in \mathbb{R}^3$ is a bounded measurable set of volume $|K|$ containing no pair of perpendicular vectors. (The latter condition is one that might reasonably be expected for the set K of momenta of a "beam" of particles.) For each $\tilde{\xi} \in \mathbb{R}^3$ and u as in (2.9) we put

$$u_{\tilde{\xi}}(\tilde{x}) = u(\tilde{x} + \tilde{\xi}). \tag{2.10}$$

We have for fixed $\tilde{\xi}$,

$$Su_{\tilde{\xi}}(\tilde{k}') = |\tilde{K}|^{-1/2}\chi_K(\tilde{k}')e^{i\tilde{k}'\cdot\tilde{\xi}} + \frac{i|K|^{-1/2}}{\pi} \times \int_K d\tilde{k}\, e^{i\tilde{k}\cdot\tilde{\xi}} f_{\tilde{k}}(\Omega)\delta(k^2 - k'^2), \tag{2.11}$$

and so,

$$\begin{aligned} P(u_{\tilde{\xi}}, C) &= \int_C d\tilde{k}'\, |Su_{\tilde{\xi}}(k')|^2 \\ &= \frac{|K \cap C|}{|K|} - \frac{2}{|K|\pi}\,\text{Im}\int_{K\cap C} d\tilde{k}' \int_K d\tilde{k}\, e^{i(\tilde{k}-\tilde{k}')\cdot\tilde{\xi}} \\ &\quad \times f_{\tilde{k}}(\Omega)\delta(k^2 - k'^2) + \frac{1}{|K|\pi^2}\int_C d\tilde{k}' \int\int_{K\times K} d\tilde{k}\, d\tilde{l} \\ &\quad \times e^{i(\tilde{k}-\tilde{l})\cdot\tilde{\xi}} f_{\tilde{k}}(\Omega) f_{\tilde{l}}(\Omega)\delta(k^2 - k'^2)\delta(l^2 - k'^2). \end{aligned} \tag{2.12}$$

We now suppose that

$$K \cap C = \emptyset. \tag{2.13}$$

Let $\tilde{k}_0 \in K$ be a fixed vector, and let $d\tilde{\xi}$ denote Lebesgue

measure on the plane defined by $\tilde{\xi}\cdot\tilde{k}_0 = 0$. We will calculate $\int_{\tilde{\xi}\cdot\tilde{k}_0} d\tilde{\xi} P(u_{\tilde{\xi}}, C)$; in doing so we will proceed rather formally in manipulations with various δ distributions; these manipulations are easily justified using standard results on Fourier transforms (see, e.g., the proof of Theorem 1 in Ref. 9). We have

$$\int_{\tilde{\xi}\cdot\tilde{\mathbf{k}}_0 = 0} d\tilde{\xi}\, e^{i(\tilde{k}-\tilde{l})\cdot\tilde{\xi}} = (2\pi)^2 \delta^2(\tilde{k}_\perp - \tilde{l}_\perp), \tag{2.14}$$

where δ^2 is the two-dimensional δ distribution and $\tilde{k}_\perp$, $\tilde{l}_\perp$ denote the components of $\tilde{k}$ and $\tilde{l}$ perpendicular to $\tilde{k}_0$. Hence

$$\begin{aligned}\int_{\tilde{\xi}\cdot\tilde{\mathbf{k}}_0 = 0} & d\tilde{\xi}\, P(u_{\tilde{\xi}}, C) \\ &= \frac{4}{|K|}\int_C d\tilde{k}' \int\int_{K\times K} d\tilde{k} d\tilde{l} f_{\tilde{k}}(\Omega)\overline{f_{\tilde{l}}(\Omega)}\delta^2(\tilde{k}_\perp - \tilde{l}_\perp) \\ &\quad \times \delta(k^2 - k'^2)\delta(l^2 - k'^2).\end{aligned} \tag{2.15}$$

We have

$$\begin{aligned}\delta^2(\tilde{k}_\perp - \tilde{l}_\perp)&\delta(k^2 - k'^2)\delta(l^2 - k'^2) \\ &= \delta^2(\tilde{k}_\perp - \tilde{l}_\perp)\delta(k^2 - l^2)\delta(l^2 - k'^2) \\ &= \delta^2(\tilde{k}_\perp - \tilde{l}_\perp)\delta(k_\parallel^2 - l_\parallel^2)\delta(l^2 - k'^2) \\ &= \frac{1}{2k_\parallel}\delta^2(\tilde{k}_\perp - \tilde{l}_\perp)\delta(k_\parallel - l_\parallel)\delta(l^2 - k'^2) \\ &= \frac{1}{2k_\parallel}\delta(\tilde{k} - \tilde{l})\delta(l^2 - k'^2),\end{aligned} \tag{2.16}$$

where $k_\parallel$ and $l_\parallel$ denote the magnitudes of the components $\tilde{k}_\parallel$ and $\tilde{l}_\parallel$ of $\tilde{k}$ and $\tilde{l}$ parallel to $\tilde{k}_0$. We have also used the identity

$$\delta(k_\parallel^2 - l_\parallel^2) = \frac{1}{2k_\parallel}[\delta(k_\parallel - l_\parallel) + (k_\parallel + l_\parallel)], \tag{2.17}$$

ane the fact that K contains no pair of perpendicular vectors to set $\delta^2(\tilde{k}_\perp - \tilde{l}_\perp)\delta(k_\parallel + l_\parallel) = 0$. (This term can only be nonzero if $\tilde{k} = \tilde{l} \perp \tilde{k}_0$.) Hence,

$$\begin{aligned}\int_{\tilde{\xi}\cdot\tilde{\mathbf{k}}_0 = 0} & d\tilde{\xi}\, P(u_{\tilde{\xi}}, C) \\ &= \frac{2}{|K|}\int_C d\tilde{k}' \int\int_{K\times K} d\tilde{k} d\tilde{l} f_{\tilde{k}}(\Omega)\overline{f_{\tilde{l}}(\Omega)}\,\frac{\delta(\tilde{k} - \tilde{l})\delta(l^2 - k'^2)}{k_\parallel} \\ &= \frac{2}{|K|}\int_C d\tilde{k}' \int_K d\tilde{k}\, |f_{\tilde{k}}(\Omega)|^2 \frac{\delta(k^2 - k')^2}{k_\parallel} \\ &= \frac{2}{|K|}\int_C d\tilde{k}' \int_K d\tilde{k}\, |f_{\tilde{k}}(\Omega)|^2 \frac{\delta(k - k')}{2kk_\parallel} \\ &= \frac{1}{|K|}\int_K d\tilde{k}\,\frac{k}{k_\parallel}\int_C d\Omega\, \sigma_{\tilde{k}}(\Omega).\end{aligned} \tag{2.18}$$

We remark that $\int_{\tilde{\xi}\cdot\tilde{\mathbf{k}}_0 = 0} d\tilde{\xi} P(u_{\tilde{\xi}}, C)$ depends on $\tilde{k}_0$ because of the term $k/k_\parallel$ in (2.18). We also note that the process of integrating over $\tilde{\xi}$ *after* computing $P(u_{\tilde{\xi}}, C)$ amounts to considering the particle as being in a *mixture* of states with different values of $\tilde{\xi}$. (Physically, $\tilde{\xi}$ corresponds to the impact parameter of the particle. See Ref. 2 for a further discussion.)

3. SCATTERING OF A BEAM OF PARTICLES

In this section we assume that K, C, and V satisfy the assumptions of the previous section which were necessary for the validity of (2.18). We now construct a multiparticle wavefunction corresponding to a system of particles each having a fairly localized momentum and with the various momenta "uniformly distributed" over K. Namely, cover $\mathbb{R}^3$ with a grid of cubes with sides parallel to the axes of length ϵ, let $K_1,\ldots,K_N$ be those intersections of K with cubes of the grid for which $|K_i| = \mathrm{vol}(K_i) > 0$ and then put

$$\psi_N(\tilde{x}_1,\ldots\tilde{x}_N) = \prod_{i=1}^N u_i(\tilde{x}_i), \tag{3.1}$$

where

$$u_i(\tilde{x}_i) = (2\pi)^{-3/2}\,|K_i|^{-1/2}\int_{K_i} d\tilde{k}\, e^{i\tilde{k}\cdot\tilde{x}_i}. \tag{3.2}$$

We have $N = O(\epsilon^{-3})$ and we shall eventually let $\epsilon\to 0$ or $N\to\infty$. As in Sec. 2 we will need to consider position translates of the particles described by the $u_i(\tilde{x}_i)$, and we put

$$\psi_{N,\tilde{\xi}} = \psi_N(\tilde{x}_1 + \tilde{\xi}_1,\ldots,\tilde{x}_N + \tilde{\xi}_N), \tag{3.3}$$

where $\tilde{\xi}_i \in \mathbb{R}^3$. We assume that the particles in $\psi_{N,\tilde{\xi}}$ interact only with the potential V (not with each other). In this case the S operator for the system of particles is just the product of the one particle S operators, and the probability $P_i(u_{i\tilde{\xi}}, C)$ that the ith particle is scattered into the cone C is given by the last term in (2.13) with K and $\tilde{\xi}$ replaced by K_i and $\tilde{\xi}_i$. (We are assuming $K\cap C = \emptyset$.) The average probability of scattering into the cone C is then given by

$$P(\psi_{N,\tilde{\xi}}, C) = \frac{1}{N}\sum_{i=1}^N P_i(u_{i,\tilde{\xi}_i}, C). \tag{3.4}$$

Now we pick some fixed $\tilde{k}_i \in K_i$ for each $i = 1,\ldots,N$; by the result of (2.18) we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N & \int_{\tilde{\xi}_i\cdot\tilde{k}_i = 0} d\tilde{\xi}_i\, P_i(u_{i,\tilde{\xi}_i}, C) \\ &= \sum_{i=1}^N \frac{1}{N|K_i|}\int_{K_i} d\tilde{k}\,\frac{k}{k_\parallel}\int_C d\Omega\, \sigma_{\tilde{k}}(\Omega),\end{aligned} \tag{3.5}$$

where in the integral over K_i, $k_\parallel$ denotes the magnitude of the component of $\tilde{k}$ parallel to $\tilde{k}_i$. We now consider the result of letting $N\to\infty$ in (3.5). Since each K_i is contained in a cube of linear dimension the order of $N^{-1/3}$ we have that the factors $k/k_\parallel$ are $1 + O(N^{-1/3})$, and so we may replace these factors by 1 in taking the limit as $N\to\infty$. Finally, if the boundary of K has (three-dimensional Lebesque) measure zero, then it is not difficult to prove that as $N\to\infty$, the right-hand side of (3.5) tends to

$$\frac{1}{|K|}\int_K d\tilde{k}\int_C d\Omega\, \sigma_{\tilde{k}}(\Omega). \tag{3.6}$$

There is another line of reasoning leading to this result which may be more satisfying than our introduction of the "average probability of scattering into C" [immediately preceeding (3.4)]. Namely, given our expression for the probability $P_i(u_{i,\tilde{\xi}_i}, C)$ that the ith particle is scattered into C, we can use this to compute the expected number of particles scattered

into C, divide by the number of particles, N, and let $N \to \infty$. This line of reasoning leads to an argument identical to the one we have presented above beginning with (3.4). In fact the expected number of particles scattered into C is given by (3.4) times N, as may be seen by arguing that this expected number is given by

$$p_1 q_2 \cdots q_n + p_2 q_1 q_3 \cdots q_n + \cdots + 2p_1 p_2 q_3 \cdots q_n + 2p_1 p_3 q_2 q_4 \cdots q_n + \cdots + 3p_1 p_2 p_3 q_4 \cdots q_n + \cdots, \tag{3.7}$$

where $p_i = P_i(u_{i,\tilde{\xi}_i}, C)$ and $q_i = 1 - p_i$, and an elementary argument shows that the sum in (3.7) is $p_1 + \cdots + p_N$ which is (3.4) times N.

We discuss briefly the interpretation of (3.6). The quantity $P(\psi_{N,\tilde{\xi}}, C)$ is for fixed $\tilde{\xi}$ a probability, but (3.5) is not, due to the $\tilde{\xi}_i$ integrations. However, for large $|\tilde{\xi}_i|$ the probability $P_i(u_{i,\tilde{\xi}_i}, C)$ is small, and so we may consider (3.6) as approximately *proportional* to the probability of scattering into C for a "wide" beam of particles with localized momenta uniformly distributed over K. The validity of (3.6) for any actual scattering experiment is dependent upon whether or not our averaging over phases (the integration over $\tilde{\xi}_i$) is well correlated with the nature of the beam of particles produced in the experiment.

[1] T.A. Green and O.E. Lanford, III, "Rigorous derivation of the phase shift formula for the Hilbert space scattering operator of a single particle," J. Math. Phys. **1**, 139–48 (1960).

[2] R.G. Newton and R. Shtokhamer, "*N*-Particle Scattering Rates," in *Physical Reality and Mathematical Description*, edited by C.P. Enz and J. Mehra (Dordrecht, The Netherlands, 1974), p. 286.

[3] J.M. Combes, R.G. Newton, and R. Shtokhamer, "Scattering into cones and flux across surfaces," Phys. Rev. D **11**, 366 (1975).

[4] J.D. Dollard, "Scattering into Cones I Potential Scattering," Commun. Math. Phys. **12**, 193–203 (1969).

[5] J.R. Taylor, *Scattering Theory: The Quantum Theory on Nonrelativistic Collisions* (Wiley, New York, 1972).

[6] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms*, Princeton Series in Physics (Princeton U.P., Princeton, New Jersey, 1971).

[7] S.T. Kuroda, "On the Existence and the Unitarity Property of the Scattering Operator," Nuovo Cimento XII, 430–54 (1959).

[8] T. Ikebe, "Eigenfunction Expansions Associated with the Schrödinger Operators and their Applications to Scattering Theory," Arch. Rat. Mech. **5**, 1–34 (1960).

[9] T. Ikebe, "On the Phase-Shift Formula for the Scattering Operator," Pac. J. Math. **15**, 511–22 (1965).

A definition of asymptotically Minkowskian space–times

S. Persides

Institute of Astronomy, University of Cambridge, United Kingdom and Department of Astronomy, University of Thessaloniki, Greece

(Received 18 April 1978)

A definition of asymptotic flatness is given without the use of Einstein's equations. For this we examine (a) the possible conditions on $\mathscr{I}^+$ and their consequences and relationships, and (b) the coordinate transformations which preserve certain forms of the unphysical metric $\tilde{\mathbf{g}}$ on a neighborhood of $\mathscr{I}^+$. Roughly, a space–time $(\mathscr{M}, \mathbf{g})$ is called *asymptotically Minkowskian at future null infinity* if it is weakly asymptotically simple, $\mathscr{I}^+$ is isometric to $\mathscr{I}^+_M$ (the $\mathscr{I}^+$ of the Minkowskian space–time) and (a) $\Omega_{;\mu\nu} = 0$, $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} = -1$, $\widetilde{C}_{\lambda\mu\rho\nu} = 0$ and $\mathscr{I}^+$ or, equivalently, (b) $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} = -1$, $R_\mu{}^\nu = O(\Omega^2)$, $\Omega_{;[\lambda}R_{\mu]}{}^\nu + \Omega R_{[\lambda}{}^\nu{}_{;\mu]} = O(\Omega^4)$ on $\mathscr{I}^+$ An equivalent definition is given based on the existence of a coordinate system in which, on a neighborhood of $\mathscr{I}^+$, $\tilde{\mathbf{g}}$ behaves as the conformal metric of Minkowski's space–time and can be written in an explicitly given form.

1. INTRODUCTION

It is generally assumed today that if general relativity or any other metric theory of gravity is correct, the space–time of a bounded source, e.g., of a binary neutron star, should behave more and more like the Minkowskian space–time as we go further away from the source. This intuitive requirement has been materialized into the concept of an "asymptotically flat space–time."

However, it seems that the concept of asymptotic flatness has not been defined rigorously and unambiguously. Since general relativity is the dominant candidate for the position of the correct gravitational theory, most often asymptotically flat space–time has been taken to mean asymptotically simple and empty space–time or, at least, a space–time for which the Ricci tensor goes fast enough to zero near $\mathscr{I}$ (the conformal boundary). This is justified by the direct relationship provided by the Einstein equations between the Ricci tensor and the energy–momentum tensor whose "vanishing at large distances" captures somehow the concept of an isolated physical system. It can be argued, however, that asymptotic flatness should be defined as a property of the space–time independently of any physical theory, e.g., without any reference to the Einstein equations, as asymptotic simplicity is defined. This attitude seems to be built into two important works by Bondi, van der Burg, and Metzner[1] and Penrose.[2–4] In the first paper Bondi and his collaborators relate asymptotic flatness to the existence of a coordinate system in which far from the source the components of the metric tensor tend somehow to the components of a flat metric. A similar approach has been adopted in more general studies.[5–8] In the second paper[2] Penrose defines asymptotic simplicity in a purely geometrical way and then goes directly to the concept of asymptotically simple and empty space–time by assuming the Einstein field equations. As a result the shear of the conformal boundary $\mathscr{I}$ vanishes, the unphysical Weyl tensor vanishes on $\mathscr{I}$, $\Omega_{;\mu\nu} = 0$ on $\mathscr{I}$ (Ω is an appropriate conformal factor) and other properties follow.[2,9] However, as it is clearly stated by Penrose,[2] the concept of asymptotically flat space–time is not defined and it is not clear which of the previous properties should be build into a definition of asymptotic flatness.

The most important argument against a condition of the form $R_\mu{}^\nu = O(\Omega^n)$ is that it cannot be tailored to have the right strength: Whatever n is, the condition is either too weak or too strong. Since most of the properties of asymptotically flat space–time are consequences[10] of $\Omega_{;\mu\nu} \hat{=} 0$ and $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ (the notation is explained at the end of this section), we would like to have a condition $R_\mu{}^\nu = O(\Omega^n)$ which will imply these two relations without imposing additional and unnecessary restrictions on the space–time. We will show in Sec. 3 that this is not possible: $R_\mu{}^\nu = O(\Omega^3)$ is not strong enough to imply $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$, while $R_\mu{}^\nu = O(\Omega^4)$ implies $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ *and* some other unnecessary restrictions for the space–time. A simple example which demonstrates the inappropriateness of the condition $R_\mu{}^\nu = O(\Omega^n)$ is given by the space–time with metric (in coordinates u, r, θ, ϕ) $= g_{00} = 1, g_{01} = 1 + r^{-2}, g_{22} = \sin^{-2}\theta g_{33} = -r^2$ and the remaining components equal to zero. With $\Omega = r^{-1}$ this space–time has (in coordinates $u, \omega = r^{-1}, \theta, \phi$) $\Omega_{;\mu\nu} \hat{=} 0$, $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ but $R_1{}^0 = -4\Omega^2 + O(\Omega^3)$. Since such a space–time should be certainly called asymptotically flat, we conclude that the condition $R_\mu{}^\nu = O(\Omega^4)$ should not be included in a definition of asymptotic flatness. Consequently a question arises: Are there any conditions containing the Ricci tensor only (and perhaps Ω) that are necessary and sufficient in order to have $\Omega_{;\mu\nu} = 0$, $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$? As it turns out (Theorems 3 and 4) such conditions do exist!

Another question related to the definition of asymptotic flatness arises from two basic theorems referring to flat ($R_{\lambda\mu\rho\nu} = 0$) and conformally flat ($C_{\lambda\mu\rho\nu} = 0$) space–time: A space–time is flat or conformally flat if, and only if, there is a coordinate system in which $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ or $g_{\mu\nu} = \Omega^{-2}\,\text{diag}(1, -1, -1, -1)$, respectively. Whatever the definition of asymptotic flatness is, it is natural, if not necessary, to ask whether a similar theorem can be formulated for an asymptotically flat space–time, that is whether there is a coordinate system in which a simple explicit form of the metric can be given satisfying automatically $\Omega_{;\mu\nu} \hat{=} 0$ and/or $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$. Again it turns out that such a system exists (Theorems 1 and 2) for space–times which satisfy the first or both of $\Omega_{;\mu\nu} \hat{=} 0$, $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$.

 0022-2488/79/081731-10$01.00

The purpose of this paper is to analyze the interrelations of the geometrical properties attributed to asymptotically flat space–times, answer the above questions, and give a reasonable, geometrical and rigorous definition of asymptotic flatness without any reference to the Einstein or other field equations. This definition will be given in three equivalent ways. The first two will be stated in terms of tensor conditions which hold only on $\mathscr{I}^+$. The third definition will be stated in terms of the existence of a coordinate system in which the metric can be written in an explicitly given form in a neighborhood of $\mathscr{I}^+$. The fact that asymptotic flatness can be defined in these three equivalent ways should be regarded as an important asset of the definition for aesthetic as well as practical reasons.

The method used in this paper to study asymptotically flat space–times consists of establishing a suitable coordinate system on $\mathscr{I}^+$ and expressing tensor relations in that particular coordinate system. This approach seems to uncover some interdependencies between tensor relations which have not been discovered by purely tensorial methods. However, it should be emphasized that all properties, theorems, and conclusions are coordinate independent and quite probably most of them can be proved by purely tensorial methods (although such proofs are not yet known). But there are also properties, e.g., Theorem 1, which can be obtained only by considering particular coordinate systems.

To avoid unnecessary restrictions on the space–time we adopt at each step only what seem to be the minimum requirements which enable us to perform the calculations. Thus we assume successively (i) existence of $\mathscr{I}^+$, (ii) appropriate internal structure for $\mathscr{I}_+$, (iii) appropriate "fastening" of $\mathscr{I}^+$ to the rest of the space–time, and (iv) other conditions on $\mathscr{I}^+$. In Sec. 2 we present some preliminary considerations regarding the Minkowski space–time, some properties of $\mathscr{I}^+$, and the basic requirements for asymptotic flatness. In Sec. 3 we examine the conditions which should be probably included in a definition of asymptotic flatness. In Sec. 4 we examine the transformations which are related to the possible behavior of the conformal metric on a neighborhood of $\mathscr{I}^+$. In Sec. 5 we present some theorems and the definition of asymptotically Minkowskian space–time. Finally in Sec. 6 we give some properties of asymptotically flat space–times and make a few concluding remarks.

In what follows the term *space–time* means as usual[11] a pair $(\mathscr{M}, \mathbf{g})$ of a connected four-dimensional Hausdorff C^∞ manifold $\mathscr{M}$ with a Lorentz metric $\mathbf{g}$ on it (i.e., a metric of signature -2). Since we restrict oursleves to signature -2, *flat* and *Minkowskian* should be regarded as equivalent terms. Other concepts will be defined when they arise. It should be pointed out, however, that intuitive and precise concepts will often appear mixed, since one of the objectives of this paper is to present arguments supporting the reasonableness of the definition of asymptotic flatness.

In this paper Latin indices will stand for 0, 2, 3, while Greek indices will stand for 0, 1, 2, 3. The components of the physical metric will be denoted by $g_{\mu\nu}$, while the components of the conformal or unphysical metric will be denoted by $\tilde{g}_{\mu\nu}$. Components of tensor quantities calculated from $\tilde{g}_{\mu\nu}$ will be distinguished by a $\sim$ sign. Covariant derivatives with respect to $\tilde{g}_{\mu\nu}$ are denoted by a semicolon. Commas denote partial derivatives. A capital letter M as superscript or subscript marks quantities referring to the Minkowskian space–time. Finally, tensor equations which hold *only on* $\mathscr{I}^+$ will be written with the special equality symbol $\hat{=}$. Such relations hold in every coordinate system, but they cannot be differentiated covariantly. At some points long but straightforward calculations are needed before a property can be established. In order to avoid unnecessary sidetracking and keep the emphasis on the important issues we have omitted such calculations and give only the results in the Appendix.

2. BASIC PROPERTIES OF $\mathscr{I}^+$

To speak about infinity of space–time we have to postulate that the space–time has an infinity. Thus it is very reasonable to assume that our space–time is somehow asymptotically simple. In this paper an orientable space–time $(\mathscr{M}, \mathbf{g})$ will be called *weakly asymptotically simple*[2,11] if there exist (a) a strongly causal space $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ with a smooth nonempty boundary $\mathscr{I}$ $(\subset \tilde{\mathscr{M}})$, (b) an open subset U of $\mathscr{M}$ and an open neighborhood $\tilde{U}$ of $\mathscr{I}$ $(\tilde{U} \subset \tilde{\mathscr{M}})$, and (c) a diffeomorphism $f: U \to \tilde{U} - \mathscr{I}$ such that there is a smooth (C^3 at least) function Ω on $\tilde{U}$ with $\Omega = 0$ and $\Omega_{;\mu} \neq 0$ on $\mathscr{I}$ and $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ on $\tilde{U} - \mathscr{I}$. This definition of asymptotic simplicity is weaker than Penrose's original definition[2] since we have omitted[12] the condition that every geodesic in $\tilde{\mathscr{M}}$ has two endpoints on $\mathscr{I}$. Thus we avoid any conditions on the interior of the space–time and define directly weak asymptotic simplicity.[13] Also we are not interested in including all infinity in $\mathscr{I}$. Finally, we asked that Ω be C^3, since later we will consider derivatives of Ω.

With this definition the first and most plausible condition to be imposed on the space–time is the following:

Condition A: *The space–time is weakly asymptotically simple.*

The most important requirement for our space–time[14] is to "behave" as the flat space–time does near $\mathscr{I}$. To exploit this qualitative statement we consider briefly the Minkowskian space–time. In null-spherical coordinates (u, r, θ, ϕ) we have $dS^2 = du^2 + 2dudr - r^2(d\theta^2 + \sin^2\theta d\phi^2)$ (u is the retarded time). Changing to coordinates $(u, \omega, \theta, \phi)$ with $\omega = r^{-1}$ we find $dS^2 = du^2 - 2\omega^{-2}dud\omega - \omega^{-2}(d\theta^2 + \sin^2\theta d\phi^2)$. Assuming a conformal factor $\Omega = \omega$ we have the metric of the conformal (unphysical) space–time

$$\tilde{b}_{\mu\nu} = \begin{bmatrix} \omega^2 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -\sin^2\theta \end{bmatrix}. \tag{1}$$

The future null infinity $\mathscr{I}^+_M$ of the Minkowskian space–time is defined[11] by the equations $\omega = 0, t = +\infty$. It is a well-known three-dimensional null hypersurface with topology $R^1 \times S^2$ and induced metric [in coordinates x^i

$= (u, \theta\phi)]$

$$\tilde{\gamma}_{ij} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -\sin^2\theta \end{bmatrix}. \tag{2}$$

From Eq. (1) it can be found that the only nonzero components of the Riemann tensor and Ricci tensor for the metric $\tilde{b}_{\mu\nu}$ are

$$\tilde{R}^{M}_{0101} = \sin^{-2}\theta \tilde{R}^{M}_{2323} = 1, \tag{3}$$

$$\tilde{R}^{M}_{00} = -\omega^2, \quad \tilde{R}^{M}_{01} = 1, \quad \tilde{R}^{M}_{22} = \sin^{-2}\theta \tilde{R}^{M}_{33} = -1, \tag{4}$$

while the Weyl tensor is obviously zero. Furthermore, we find easily that

$$\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} = -1, \quad \Omega^{-1}\Omega_{;\mu}{}^{\mu} = -2. \tag{5}$$

To make our general space–time behave a little more like the Minkowskian space–time we ask that the two space–times have identical conformal boundaries at least partially. Thus we impose the second condition:

Condition B: There is a subset $\mathscr{I}^+$ of $\mathscr{I}$ which is isometric to $\mathscr{I}^+_M$.

Thus $\mathscr{I}^+$ inherits all the *intrinsic* properties of $\mathscr{I}^+_M$, that is its shape, its size, the possible coordinate systems, etc.[15] It seems reasonable to build Conditions A and B into a definition of asymptotic flatness. Furthermore they cannot be made weaker and any replacing of them, e.g., replacing condition B by "$R_{\mu}{}^{\nu} \to 0$ fast enough near $\mathscr{I}^+$," imposes much heavier restrictions on the space–time. However, they guarantee only the existence of $\mathscr{I}^+$ and its similarity to $\mathscr{I}^+_M$. They do not tell whether $\mathscr{I}$ is essentially larger than $\mathscr{I}^+$, whether $\mathscr{I}^-$ exists, how $\mathscr{I}^+$ is tied up with the rest of the space–time, i.e., "how much" the space–time resembles the Minkowksi space–time near $\mathscr{I}^+$. This "degree of resemblance" is the crucial missing ingredient for a definition of asymptotic flatness. Should we demand that Eqs. (3), (4), (5), $\Omega_{;\mu\nu} \hat{=} 0$, $\tilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$, etc. be satisfied? To see how far it is reasonable to go in demanding Minkowskian behavior near $\mathscr{I}^+$ we consider the possible behavior of the metric tensor on a neighborhood of $\mathscr{I}^+$.

Let $x^i = (u, \theta, \phi)$ be a coordinate system for $\mathscr{I}^+$ and $x^{\mu} = (u, \omega, \theta, \phi)$ a coordinate system on a neighborhood $\tilde{U}$ of $\mathscr{I}^+$ with $\omega = 0$ on $\mathscr{I}^+$. Since $\tilde{\mathbf{g}}$ is assumed at least C^3 on $\mathscr{I}^+$ we can write

$$\tilde{g}_{\mu\nu} = \alpha_{\mu\nu} + \beta_{\mu\nu}\omega + \gamma_{\mu\nu}\omega^2 + \delta_{\mu\nu}\omega^3 + O_4 \tag{6}$$

on $\tilde{U}$. In this notation $O_n = O(\omega^n)$ stands for terms which together with their derivatives with respect to x^i go to zero faster than $\omega^{n-\epsilon}$, while their derivative with respect to ω goes to zero faster than $\omega^{n-1-\epsilon}$ when $\omega \to 0$. Since $\mathscr{I}^+$ is isometric to $\mathscr{I}^+_M$ we can choose the coordinate system so that

$$\tilde{g}_{ij} \hat{=} \tilde{\gamma}_{ij}. \tag{7}$$

Hence we have

$$\alpha_{ij} = 0 \quad \text{except} \quad \alpha_{22} = \sin^{-2}\theta\alpha_{33} = -1. \tag{8}$$

Changing the scale of u on $\mathscr{I}^+$ we can make $\alpha_{01} = -1$ ($\alpha_{01} \neq 0$ since $\det\tilde{g}_{\mu\nu} \neq 0$ on $\mathscr{I}^+$). Thus the most general form for $\tilde{g}_{\mu\nu}$ on $\tilde{U}$ is

$$\tilde{g}^{(1)}_{\mu\nu} = \begin{bmatrix} O_1 & -1+O_1 & O_1 & O_1 \\ -1+O_1 & O_0 & O_0 & O_0 \\ O_1 & O_0 & -1+O_1 & O_1 \\ O_1 & O_0 & O_1 & -\sin^2\theta + O_1 \end{bmatrix} \tag{9}$$

which is a special case of Eq. (6). Setting more $\alpha_{\mu\nu}, \beta_{\mu\nu}$, etc. equal to zero we can further specialize the metric. To examine the simplest possible form of $\tilde{g}_{\mu\nu}$ we will need four more forms. Let us define

$$\tilde{g}^{(2)}_{\mu\nu} = \tilde{g}^{(1)}_{\mu\nu} \quad \text{with} \quad \beta_{00} = 0, \tag{10}$$

$$\tilde{g}^{(3)}_{\mu\nu} = \tilde{g}^{(2)}_{\mu\nu} \quad \text{with} \quad \alpha_{11} = \alpha_{12} = \alpha_{13} = \beta_{01} = \beta_{02} = \beta_{03} = 0, \tag{11}$$

$$\tilde{g}^{(4)}_{\mu\nu} = \tilde{g}^{(3)}_{\mu\nu} \quad \text{with} \quad \gamma_{00} = 1. \tag{12}$$

In general there is no transformation which will reduce $\tilde{g}^{(2)}_{\mu\nu}$ to $\tilde{g}^{(3)}_{\mu\nu}$ or $\tilde{g}^{(3)}_{\mu\nu}$ to $\tilde{g}^{(4)}_{\mu\nu}$. This will be proved in Sec. 4, where we will examine the additional conditions the space–time has to satisfy for these reductions to be possible. It should be noted that from $\tilde{g}_{\mu\nu}$ to $\tilde{g}^{(1)}_{\mu\nu}$, $\tilde{g}^{(2)}_{\mu\nu}$, etc. the metric of the space–time looks more and more like the conformal Minkowskian metric $\tilde{b}_{\mu\nu}$.

3. CONDITIONS ON $\mathscr{I}^+$

The candidates we will consider as additional conditions on the space–time can be selected from the covariant properties of asymptotically simple and empty spacetime and perhaps some properties of the Minkowski space–time.

In what follows we examine each one of these conditions, their relationships and how they affect the metric. An evaluation of the consequences of each condition will indicate which conditions should be included in a definition of asympotic flatness.

(i) The condition $\Omega_{;\mu}\Omega^{;\mu} \hat{=} 0$: Since $\Omega = \omega$ we have $\Omega_{;\mu} = \omega_{;\mu} = [0, 1, 0, 0]$ and $\Omega_{;\mu}\Omega^{;\mu} = \tilde{g}^{\mu\nu}\Omega_{;\mu}\Omega_{;\nu} = \tilde{g}^{11}$ which vanishes identically on $\mathscr{I}^+$ for $\tilde{g}^{(1)}_{\mu\nu}$ and consequently all the other forms of the metric.

(ii) The condition $\Omega_{;\mu}{}^{\mu} \hat{=} 0$: We have $\Omega_{;\mu}{}^{\mu} = \tilde{g}^{\mu\nu}\Omega_{;\mu\nu} = -\tilde{g}^{\mu\nu}\tilde{\Gamma}^{1}_{\mu\nu} \hat{=} -\beta_{00}$. For $\tilde{g}^{(1)}_{\mu\nu}$ this condition implies

$$\beta_{00} = 0, \tag{13}$$

while it is satisfied automatically by $\tilde{g}^{(2)}_{\mu\nu}$, $\tilde{g}^{(3)}_{\mu\nu}$ and $\tilde{g}^{(4)}_{\mu\nu}$.

(iii) The condition $\Omega^{-1}\Omega^{\mu}_{;\mu} \hat{=} -2$: The implications of this condition on $\tilde{g}^{(1)}_{\mu\nu}$ and $\tilde{g}^{(2)}_{\mu\nu}$ are complicated and, as it turns out, they are not needed. For $\tilde{g}^{(3)}_{\mu\nu}$ the condition implies (for the notation see the Appendix)

$$\Phi_{,0} - 2\gamma_{00} + 2 = 0 \tag{14}$$

while for $\tilde{g}^{(4)}_{\mu\nu}$ gives

$$\Phi_{,0} = 0. \tag{15}$$

(iv) The condition $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$: We have

$\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} = \omega^{-2}\tilde{g}^{11} = \beta^{11}\omega^{-1} + \gamma^{11} + O_1$. [$\beta^{\mu\nu}$, $\gamma^{\mu\nu}$, etc. are the coefficients of $\tilde{g}^{\mu\nu}$ in an equation similar to Eq. (6).] For $\tilde{g}^{(1)}_{\mu\nu}$ this condition implies $\beta^{11} = 0$, $\gamma^{11} = -1$ which are equivalent to

$$\beta_{00} = 0, \quad \gamma_{00} + \beta_{02}{}^2 + \sin^{-2}\theta\beta_{03}{}^2 = 1. \tag{16}$$

For $\tilde{g}^{(2)}_{\mu\nu}$ the condition implies only the second of Eqs. (16). For $\tilde{g}^{(3)}_{\mu\nu}$ the condition implies

$$\gamma_{00} = 1, \tag{17}$$

while it is satisfied by $\tilde{g}^{(4)}_{\mu\nu}$.

(v) The condition $u_{;\mu}u^{;\mu} \hat{=} 0$, $\theta_{;\mu}\theta^{;\mu} \hat{=} -1$, $\theta_{;\mu}\phi^{;\mu} \hat{=} 0$, etcetera: Obviously these conditions are satisfied (e.g., $\theta_{;\mu}\theta^{;\mu} \hat{=} -1$) or imply some restrictions (e.g., $u_{;\mu}u^{;\mu} \hat{=} \alpha^{00} = 0$) for the metric. However, they have an essential disadvantage: To write them down we have to assume the existence of some new scalar functions u, θ, ϕ, which means esentially that a special coordinate system exists. Such conditions should not be included in a coordinate independent definition of asymptotic flatness.

(vi) The condition[16] $\Omega_{;\mu\nu} - \frac{1}{4}\tilde{g}_{\mu\nu}\Omega_{;\rho}{}^{\rho} \hat{=} 0$: In our coordinate system this condition gives

$$\Omega_{;\mu\nu} - \tfrac{1}{4}\tilde{g}_{\mu\nu}\Omega_{;\rho}{}^{\rho} \hat{=} -\widetilde{\Gamma}^{1}_{\mu\nu} + \tfrac{1}{4}\beta_{00}\alpha_{\mu\nu} \hat{=} 0. \tag{18}$$

Written explicitly this equation gives

$$\beta_{00} = \beta_{01} - \tfrac{1}{2}\alpha_{11,0} = \beta_{02} - \alpha_{12.0} = \beta_{03} - \alpha_{13,0} = 0. \tag{19}$$

These are restrictions for $\tilde{g}^{(1)}_{\mu\nu}$ and $\tilde{g}^{(2)}_{\mu\nu}$, while they are satisfied automatically by $\tilde{g}^{(3)}_{\mu\nu}$ and $\tilde{g}^{(4)}_{\mu\nu}$.

(vii) The condition $\sigma_{\mu\nu} \hat{=} 0$: The shear of the congruence of the curves along which u, θ, ϕ are constants is

$$\sigma_{\mu\nu} = \widetilde{P}_{\lambda(\mu}\Omega_{;\nu)}{}^{\lambda} - \tfrac{1}{3}\Omega_{;\lambda}{}^{\lambda}\widetilde{P}_{\mu\nu}, \tag{20}$$

where

$$\widetilde{P}_{\mu\nu} = \tilde{g}_{\mu\nu} + \Omega_{;\mu}\Omega_{;\nu}. \tag{21}$$

Evaluating $\sigma_{\mu\nu}$ in our coordinate system we find

$$\sigma_{\mu\nu} \hat{=} -\widetilde{\Gamma}^{1}_{\mu\nu} + \tfrac{1}{6}\beta_{00}(2\alpha_{\mu\nu} - \delta^{1}_{\mu}\delta^{1}_{\nu}). \tag{22}$$

Setting this quantity equal to zero we find again equations (19). Hence this condition is equivalent to the previous one.

(viii) The condition $\Omega_{;\mu\nu} \hat{=} 0$: In our coordinate system we have $\Omega_{;\mu\nu} = -\widetilde{\Gamma}^{1}_{\mu\nu}$. Setting this quantity equal to zero we find again Eqs. (19). Hence this condition is equivalent to condition (vi) as well as to condition (vii).

(ix) The condition $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$: To explicity write this condition we evaluate the components of the Riemann and Ricci tensors from the unphysical metric in our coordinate system. After some long but straightforward calculations we find that the restrictions imposed on the metric can be given by two sets of equations. The first set results from the equations $\widetilde{C}_{0102} \hat{=} \widetilde{C}_{0103} \hat{=} \widetilde{C}_{0123} \hat{=} 0$ which give $\widetilde{R}_{0102} \hat{=} \widetilde{R}_{0103} \hat{=} \widetilde{R}_{0123} \hat{=} 0$ or

$$\alpha_{12,00} + \beta_{00,2} - \beta_{02,0} = 0, \tag{23}$$

$$\alpha_{13,00} + \beta_{00,3} - \beta_{03,0} = 0, \tag{24}$$

$$\alpha_{13,02} - \alpha_{12.03} + \beta_{02,3} - \beta_{03,2} = 0. \tag{25}$$

The second set of equations is obtained from $\widetilde{C}_{0101} \hat{=} \widetilde{C}_{0112} \hat{=} \widetilde{C}_{0113} \hat{=} \widetilde{C}_{1212} \hat{=} \widetilde{C}_{1213} \hat{=} 0$. Linear combinations of these equations give an equivalent set of equations with left-hand sides equal to $\widetilde{R}_{0101}$, $\widetilde{R}_{0112}$, $\widetilde{R}_{0113}$, $\widetilde{R}_{1212} - \sin^{-2}\theta R_{1313}$, $\widetilde{R}_{1213}$, and right-hand sides containing only $\alpha_{\mu\nu}$ and $\beta_{\mu\nu}$ (not $\gamma_{\mu\nu}$). The remaining $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ give identities.

There are some essential differences between the two sets of equations. The first set contains some $\alpha_{\mu\nu}$ and $\beta_{\mu\nu}$ but no $\gamma_{\mu\nu}$. The equations of the second set contain $\gamma_{\mu\nu}$ in a characteristic way. Since each component of $\widetilde{R}_{\lambda\mu\rho\nu}$ is evaluated on $\mathscr{I}^{+}$, that is for $\omega = 0$, the only case where a $\gamma_{\mu\nu}$ will appear in these equations is when $\widetilde{R}_{\lambda\mu\rho\nu}$ has *two* indices equal to 1 so that the second derivative with respect to ω will generate from $\widetilde{R}_{1\mu1\nu}$ the term $\gamma_{\mu\nu}$ with coefficient ± 1. Hence the equations of the second set can be solved and give explicitly γ_{00}, γ_{02}, γ_{03}, $\gamma_{22} - \sin^{-2}\theta\gamma_{33}$, γ_{23} in terms of $\alpha_{\mu\nu}$ and $\beta_{\mu\nu}$. Thus $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ imposes severe restrictions on $\gamma_{\mu\nu}$, that is the third term in the expansion of $\tilde{g}_{\mu\nu}$ in powers of ω. Hence the first set of equations imposes restriction on $\tilde{g}^{(1)}_{\mu\nu}$ and $\tilde{g}^{(2)}_{\mu\nu}$, while it is satisfied automatically by $\tilde{g}^{(3)}_{\mu\nu}$ and $\tilde{g}^{(4)}_{\mu\nu}$. The second set imposes conditions on all forms of $\tilde{g}_{\mu\nu}$. It should be also noted that the equations of the first set are consequences of condition (vi) or (vii) or (viii), while the equations of the second set cannot be related to any previous conditions.

If Eqs. (19) are satisfied, then according to Theorem 1 (Sec. 5) there is a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(3)}_{\mu\nu}$. In this coordinate system $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ gives only the second set of equations which can be written (see the Appendix)

$$A = B = C = E = F = 0. \tag{26}$$

These equations will be used to study the relationships between the conditions $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ and $R_{\mu}{}^{\nu} = O(\Omega^{n})$.

(x) The condition $R_{\mu}{}^{\nu} \hat{=} O(\Omega^{n})$: (a) Let $n = 2$. Then $\Omega^{-1}R_{\mu}{}^{\nu} \hat{=} 0$ and a straightforward calculation gives Eqs. (19). Consequently the condition $R_{\mu}{}^{\nu} = O(\Omega^{2})$ is equivlaent to condition (vi) or (vii) or (viii).

(b) Let $n = 3$. Then $\Omega_{;\mu\nu} \hat{=} 0$ and according to Theorem 1 (Sec. 5) there is a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(3)}_{\mu\nu}$. In this coordinate system the condition $\Omega^{-2}R_{\mu}{}^{\nu} \hat{=} 0$ gives (see the Appendix) the equivalent set

$$A = B = C = D = 0. \tag{27}$$

Hence the condition $R_{\mu}{}^{\nu} = O(\Omega^{3})$ implies *only* the first three of Eqs. (26). It does not imply $E = F = 0$, while also gives $D = 0$, which is a completely unnecessary requirement for asymptotic flatness.

(c) Let $n = 4$. Then again in a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(3)}_{\mu\nu}$ we obtain Eqs. (27) and the additional relations

$$E_{,0} = F_{,0} = 0, \tag{28}$$

$$E_{,2} + 2\cot\theta E + \sin^{-2}\theta F_{,3} = 0, \tag{29}$$

$$E_{,3} - F_{,2} - \cot\theta F = 0, \tag{30}$$

$$G = 0. \tag{31}$$

From Eqs. (28)–(30) we conclude that $\sin^2\theta\cdot E$ and $\sin\theta\cdot F$ are independent of u and satisfy the equation

$$(Y_{,2}\sin\theta)_{,2}\sin\theta + Y_{,33} = 0. \tag{32}$$

This equation coupled with the requirement that E and F must be regular on the unit sphere implies

$$E = F = 0. \tag{33}$$

Hence, as is well known,[2,9,14] $R_\mu{}^\nu = O(\Omega^4)$ implies Eqs. (26), that is $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$. But it also imposes on the space–time the conditions $D = G = 0$. The equation $G = 0$ gives the first condition containing[17] $\delta_{\mu\nu}$. Obviously there is no reason to impose such severe and unnecessary restrictions on the space–time.

Hence we conclude that $R_\mu{}^\nu = O(\Omega^n)$ is either too weak or too strong in order to have $\Omega_{;\mu\nu} \hat{=} 0$ and $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$.

(xi) The condition $\widetilde{R} \hat{=} 0$: For $\tilde{g}^{(1)}_{\mu\nu}$ and $\tilde{g}^{(2)}_{\mu\nu}$ this condition is complicated and it is not needed. For $\tilde{g}^{(3)}_{\mu\nu}$ it gives

$$2\Phi_{,0} - \gamma_{00} + 1 = 0, \tag{34}$$

while for $\tilde{g}^{(4)}_{\mu\nu}$ it again gives Eq. (15).

(xii) The condition $\widetilde{R}_{\mu\nu} \hat{=} \widetilde{R}^M_{\mu\nu}$: This condition gives five relations containing $\alpha_{\mu\nu}$ and $\beta_{\mu\nu}$ only and four also containing some $\gamma_{\mu\nu}$. However, these restrictions on the metric are too severe and do not hold for metrics which satisfy the Einstein equations.

(xiii) The condition $\widetilde{R}_{\lambda\mu\rho\nu} \hat{=} \widetilde{R}^M_{\lambda\mu\rho\nu}$: The same remarks apply in this case as in the previous one. This condition implies (ix), (xi) and (xii).

The following additional remarks can be made from the study of conditions (i)–(xiii): (a) Conditions (vi)–(viii) and (xa) are equivalent for any conformal factor Ω and independently of any field equations. They cannot be derived from any other condition.

(b) Condition (iv) is the only one which in the appropriate coordinate system gives $\gamma_{00} = 1$, namely it makes $\tilde{g}_{00}$ start with an ω^2 term (as $\tilde{b}_{\mu\nu}$ does).

(c) If conditions (iv) and (viii) hold, then condition (iii) is equivalent to condition (xi) and condition (ix) implies (iii) and (xi).

On the basis of the previous considerations it seems that, if we want the space–time to have Minkowskian-like metric near $\mathscr{I}^+$, we can impose either conditions (iv) and (viii) *or* (iv), (viii), and (ix) depending on how much we want to restrict the space–time. Similar conclusions will be reached in the next section from a study of the possible transformations.

4. TRANSFORMATIONS

Studies of the asymptotic symmetries of an asymptotically flat space–time have been carried out[1,5,18] on the physical space–time $(\mathscr{M}, \mathbf{g})$ in coordinates (u, r, θ, ϕ) and have resulted in establishing an important group of transformations, the BMS group. It can be argued however, that it is more appropriate to study the asymptotic symmetries of a space–time on a neighborhood of $\mathscr{I}^+$ (which includes $\mathscr{I}^+$) and this is possible only on the conformal space–time $(\widetilde{\mathscr{M}}, \tilde{g})$.

In this section we examine the transformations related to the conformal space–time $(\widetilde{\mathscr{M}}, \tilde{g})$ in coordinates $(u, \omega, \theta, \phi)$ on a neighborhood of $\mathscr{I}^+$. Specifically, there are two important questions related to the behavior of $\tilde{\mathbf{g}}$ on a neighborhood of $\mathscr{I}^+$. First, we can ask what transformations preserve the form of $\tilde{g}^{(1)}_{\mu\nu}$ or $\tilde{g}^{(2)}_{\mu\nu}$, etc. Second, we can ask whether or under what conditions $\tilde{g}^{(1)}_{\mu\nu}$ can be reduced to $\tilde{g}^{(2)}_{\mu\nu}$, $\tilde{g}^{(2)}_{\mu\nu}$ to $\tilde{g}^{(3)}_{\mu\nu}$, etc. The properties found exhibit many similarities with previous results but also some striking differences which are due to the fact that a conformal factor Ω has been used in going from $(\mathscr{M}, \mathbf{g})$ to $(\widetilde{\mathscr{M}}, \tilde{\mathbf{g}})$.

In what follows we will repeatedly need a transformation of the unit sphere onto itself. To simplify the statements and notation we will denote by $\{\Theta,\Phi\} = R\{\Theta',\Phi'\}$ a transformation $(\Theta,\Phi) \to (\Theta',\Phi')$ such that

$$d\Theta^2 + \sin^2\Theta d\Phi = d\Theta'^2 + \sin^2\Theta' d\Phi' \tag{35}$$

for $0 \leqslant \Theta,\Theta' \leqslant \pi$ and $0 \leqslant \Phi,\Phi' \leqslant 2\pi$ (the points $\Phi = 0$ are identified with the points $\Phi = 2\pi$). The necessary and sufficient conditions for a transformation $(\Theta,\Phi) \to (\Theta',\Phi')$ to be of this kind are

$$\left(\frac{\partial\Theta}{\partial\Theta'}\right)^2 + \sin^2\Theta\left(\frac{\partial\Phi}{\partial\Theta'}\right)^2 = 1, \tag{36}$$

$$\left(\frac{\partial\Theta}{\partial\Phi'}\right)^2 + \sin^2\Theta\left(\frac{\partial\Phi}{\partial\Phi'}\right)^2 = \sin^2\Theta', \tag{37}$$

$$\frac{\partial\Theta}{\partial\Theta'}\frac{\partial\Theta}{\partial\Phi'} + \sin^2\Theta\frac{\partial\Phi}{\partial\Theta'}\frac{\partial\Phi}{\partial\Phi'} = 0. \tag{38}$$

The Jacobian $J = J(\Theta,\Phi;\Theta',\Phi')$ of such a transformation $(\Theta,\Phi) \to (\Theta',\Phi')$ satisfies the equation

$$J^2 = \sin^2\Theta \sin^{-2}\Theta'. \tag{39}$$

It is also assumed that $\{\Theta,\Phi\} = R\{\Theta',\Phi'\}$ means that Θ and Φ may depend only on Θ',Φ', i.e., they are independent of any other parameter or variable.

Now let $(u,\omega,\theta,\phi) \to (u',\omega',\theta',\phi')$ be a transformation which is smooth (at least C^3) on $\mathscr{I}^+$ with the hypersurface $\mathscr{I}^+$ given by $\omega = 0$ or $\omega' = 0$. Such a transformation can be written in the form

$$u = u_0 + u_1\omega' + u_2\omega'^2 + O_3, \tag{40}$$

$$\omega = \omega_1\omega' + \omega_2\omega'^2 + O_3, \tag{41}$$

$$\theta = \theta_0 + \theta_1\omega' + \theta_2\omega'^2 + O_3, \tag{42}$$

$$\phi = \phi_0 + \phi_1\omega' + \phi_2\omega'^2 + O_3 \tag{43}$$

on a neighborhood of $\mathscr{I}^+$ where u_0, θ_0, ϕ_0, u_1, ω_1, etc., are functions of u', θ', ϕ'. Note that because of Eq. (41) a function which is $O(\omega^n)$ is also $O(\omega'^n)$ and is denoted by O_n.

The calculation of the components $\tilde{g}'^{(1)}_{\mu\nu}$ is straightforward. The condition $\tilde{g}'^{(1)}_{00} = O_1$ gives $\partial\theta_0/\partial u' = \partial\phi_0/\partial u' = 0$. Hence $\theta_0 = \theta_0(\theta',\phi')$ and $\phi_0 = \phi_0(\theta',\phi')$. Then the conditions $\tilde{g}'^{(1)}_{22} = -1 + O_1$,

$\tilde{g}'^{(1)}_{33} = -\sin^2\theta' + O_1$ and $\tilde{g}'^{(1)}_{23} = O_1$ give Eqs. (36)–(38) with $\Theta, \Phi, \Theta', \Phi'$ replaced by $\theta_0, \phi_0, \theta', \phi'$, respectively. Finally, from $\tilde{g}'^{(1)}_{01} = -1 + O_1$ we have $\omega_1(\partial u_0/\partial u') = 1$. The remaining $\tilde{g}'^{(1)}_{\mu\nu}$ give no additional restrictions. Thus we have the following proposition:

Proposition 1: The most general transformation which preserves the form of $\tilde{g}^{(1)}_{\mu\nu}$ is given by Eqs. (40)–(43) with

$$\{\theta_0,\phi_0\} = R\{\theta',\phi'\}, \tag{44}$$

$$\omega_1 \frac{\partial u_0}{\partial u'} = 1. \tag{45}$$

To preserve the form of $\tilde{g}^{(2)}_{\mu\nu}$ we add the additional restriction that $\tilde{g}'^{(2)}_{00} = O_2$. Then we find $\partial\omega_1/\partial u' = 0$. Hence we have the following:

Proposition 2: The most general transformation which preserves the form of $\tilde{g}^{(2)}_{\mu\nu}$ is given by Eqs. (40)–(43) with

$$\{\theta_0,\phi_0\} = R\{\theta',\phi'\}, \tag{46}$$

$$u_0 = \frac{1}{\omega_1}(u' + f), \tag{47}$$

where ω_1 and f are arbitrary functions of θ' and ϕ' only.

The transformation (46)–(47) is the closest we will find to a BMS tranformation. The difference is that (46) has a conformal factor equal to 1.

To preserve the form of $\tilde{g}^{(3)}_{\mu\nu}$ we have to ask that $\tilde{g}'^{(3)}_{01} = -1 + O_2, \tilde{g}'^{(3)}_{02}, \tilde{g}'^{(3)}_{03}$ be O_2 and $\tilde{g}'^{(3)}_{11}, \tilde{g}'^{(3)}_{12}, \tilde{g}'^{(3)}_{13}$ be O_1. In addition to Eqs. (46) and (47) we find that the transformation (40)–(43) should satisfy the equations

$$\theta_1^2 + \sin^2\theta_0\phi_1^2 + 2u_1\omega_1 = 0, \tag{48}$$

$$\frac{\partial\theta_0}{\partial\theta'}\theta_1 + \sin^2\theta_0\frac{\partial\phi_0}{\partial\theta'}\phi_1 + \frac{\partial u_0}{\partial\theta'}\omega_1 = 0, \tag{49}$$

$$\frac{\partial\theta_0}{\partial\phi'}\theta_1 + \sin^2\theta_0\frac{\partial\phi_0}{\partial\phi'}\phi_1 + \frac{\partial u_0}{\partial\phi'}\omega_1 = 0, \tag{50}$$

$$\theta_1\frac{\partial\theta_1}{\partial u'} + \sin^2\theta_0\,\phi_1\frac{\partial\phi_1}{\partial u'} + \omega_1\frac{\partial u_1}{\partial u'} + 2\omega_2\frac{\partial u_0}{\partial u'} = 0, \tag{51}$$

$$\frac{\partial\theta_0}{\partial\theta'}\frac{\partial\theta_1}{\partial u'} + \sin^2\theta_0\frac{\partial\phi_0}{\partial\theta'}\frac{\partial\phi_1}{\partial u'} + \frac{\partial u_0}{\partial u'}\frac{\partial\omega_1}{\partial\theta'} = 0, \tag{52}$$

$$\frac{\partial\theta_0}{\partial\phi'}\frac{\partial\theta_1}{\partial u'} + \sin^2\theta_0\frac{\partial\phi_0}{\partial\phi'}\frac{\partial\phi_1}{\partial u'} + \frac{\partial u_0}{\partial u'}\frac{\partial\omega_1}{\partial\phi'} = 0. \tag{53}$$

Differentiating (49) and (50) with respect to u' and subtracting from (52) and (53) respectively we find $\partial\omega_1/\partial\theta' = \partial\omega_1/\partial\phi' = 0$. Differentiating (48) with respect to u' and subtracting from (51) we find $\omega_2 = 0$. Finally we solve (49) and (50) with respect to θ_1 and ϕ_1. Thus we have the following proposition.

Proposition 3: The most general transformation which preserves the form of $\tilde{g}^{(3)}_{\mu\nu}$ is given by Eqs. (40)–(43) with

$$\{\theta_0,\phi_0\} = R\{\theta',\phi'\}, \quad u_0 = \frac{1}{\omega_1}[u' + f(\theta',\phi')], \tag{54}$$

$$\theta_1 = -\frac{J(f,\phi_0)}{J(\theta_0,\phi_0)}, \quad \phi_1 = \frac{1}{\sin^2\theta_0}\frac{J(f,\theta_0)}{J(\theta_0,\phi_0)}, \tag{55}$$

$$\omega_1 = \text{const} \neq 0, \quad \omega_2 = 0, \tag{56}$$

$$u_1 = -\frac{\omega_1}{2}\left[\left(\frac{\partial f}{\partial\theta'}\right)^2 + \frac{1}{\sin^2\theta'}\left(\frac{\partial f}{\partial\phi'}\right)^2\right], \tag{57}$$

where $J(\Theta,\Phi)$ is the Jacobian $J(\Theta,\Phi;\theta',\phi')$.

Note that in the above transformation θ_0, ϕ_0, f, u_1, θ_1, and ϕ_1 depend only on θ', ϕ', while ω_1 is independent of u', θ', ϕ'.

Finally, to find the transformation which preserves the form of $\tilde{g}^{(4)}_{\mu\nu}$ we add the condition $\tilde{g}'^{(4)}_{00} = \omega'^2 + O_3$. We find no additional restriction for the transformation. Hence we have the following:

Proposition 4: The most general transformation which preserves the form of $\tilde{g}^{(4)}_{\mu\nu}$ is given by Eqs. (40)–(43) with Eqs. (54)–(57) satisfied.

It should be emphasized that the coefficients of ω' (e.g., u_2, θ_2, etc.) which do not appear in each case are completely arbitrary. Also the transformation of Proposition 3 or 4 contains the transformation $\{\theta_0,\phi_0\} = R\{\theta',\phi'\}$, $u_0 = u' + f(\theta',\phi')$ which is a BMS transformation with conformal factor $K = 1$. If T_1, T_2, T_3, T_4 are the classes of transformations which preserve $\tilde{g}^{(1)}_{\mu\nu}, \tilde{g}^{(2)}_{\mu\nu}, \tilde{g}^{(3)}_{\mu\nu}, \tilde{g}^{(4)}_{\mu\nu}$ respectively (that is the transformations of Propositions 1, 2, 3, 4), then the following can be easily proven:

Proposition 5: Each one of T_1, T_2, T_3, T_4 is a group and $T_1 \supset T_2 \supset T_3 = T_4$.

We turn now to the second general question: Are there transformations which reduce one form of $\tilde{g}_{\mu\nu}$ to another? To reduce $\tilde{g}^{(1)}_{\mu\nu}$ to $\tilde{g}^{(2)}_{\mu\nu}$ we must find a transformation which satisfies Eqs. (44) and (45) and gives $\tilde{g}'^{(1)}_{\mu\nu} = O_2$. This condition gives $\partial\omega_1/\partial u' = \frac{1}{2}\beta_{00}$ which does not contradict with Eqs. (44) and (45). Hence there is such a transformation (but it depends on the metric). To reduce $\tilde{g}^{(2)}_{\mu\nu}$ to $\tilde{g}^{(3)}_{\mu\nu}$ we must find a transformation which satisfies Eqs. (46) and (47) and six more equations with left-hand sides those of Eqs. (48)–(53) but with right-hand sides linear combinations of $\beta_{01}, \beta_{02}, \beta_{03}, \alpha_{11}, \alpha_{12}, \alpha_{13}$. After some calculations it can be shown that this system has no solution in general. Hence in general $\tilde{g}^{(2)}_{\mu\nu}$ *cannot* be reduced to $\tilde{g}^{(3)}_{\mu\nu}$. Finally, to reduce $\tilde{g}^{(3)}_{\mu\nu}$ to $\tilde{g}^{(4)}_{\mu\nu}$ we must find a transformation which satisfies Eqs. (54)–(57) and gives $\tilde{g}'^{(3)}_{00} = \omega'^2 + O_3$. This condition gives $\gamma_{00} = 1$ which is not true in general. The above results are summarized in the following proposition.

Proposition 6: There are transformations which will reduce $\tilde{g}^{(1)}_{\mu\nu}$ to $\tilde{g}^{(2)}_{\mu\nu}$. In general there are no transformations which will reduce $\tilde{g}^{(2)}_{\mu\nu}$ to $\tilde{g}^{(3)}_{\mu\nu}$ or $\tilde{g}^{(3)}_{\mu\nu}$ to $\tilde{g}^{(4)}_{\mu\nu}$.

We conclude that if we want an asymptotically Minkowskian space–time to behave as $\tilde{g}^{(3)}_{\mu\nu}$ or $\tilde{g}^{(4)}_{\mu\nu}$ in some appropriate coordinate system, this property must be imposed on the space–time by a coordinate-independent condition. Hence we have the same choices of appropriate conditions as in Sec. 3. We have only to find out which conditions will guarantee the existence of a coordinate system in which the metric will have the form of $\tilde{g}^{(3)}_{\mu\nu}$ or $\tilde{g}^{(4)}_{\mu\nu}$.

If $\tilde{g}_{\mu\nu} = \tilde{g}^{(4)}_{\mu\nu}$ we can easily prove that there is a unique

choice of u_2, θ_2, ϕ_2 such that the transformation (40)–(43) gives $\tilde{g}'_{11}, \tilde{g}'_{12}, \tilde{g}'_{13}$ all O_2. Hence we have the following proposition which will be used to simplify $\tilde{g}_{\mu\nu}$ in the definition of Sec. 5:

Proposition 7: If $\tilde{g}_{\mu\nu} = \tilde{g}^{(4)}_{\mu\nu}$, then there is a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(5)}_{\mu\nu}$, where $\tilde{g}^{(5)}_{\mu\nu} = \tilde{g}^{(4)}_{\mu\nu}$ but with $\beta_{11} = \beta_{12} = \beta_{13} = 0$.

5. THEOREMS AND DEFINITION

In this section we give two theorems which relate the conditions of Sec. 3 to the existence of a coordinate system in which the metric takes a specific form and two theorems which essentially give the necessary and sufficient conditions on the Ricci tensor of the physical space–time in order to have $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$. These theorems indicate which space–time should be called asymptotically flat in order to have a definition in a covariant way as well as in terms of a special coordinate system.

Theorem 1: Let a space–time $(\mathscr{M}, \mathbf{g})$ be weakly asymptotically simple at future null infinity and $(\widetilde{\mathscr{M}}, \tilde{\mathbf{g}})$ the conformal space with conformal factor Ω and (future null) boundary $\mathscr{I}^+$ defined by $\Omega = 0$. Then the hypersurface $\mathscr{I}^+$ is isometric to $\mathscr{I}^+_M$ and $\Omega_{;\mu\nu} \hat{=} 0$ if and only if there is a coordinate system $(u, \omega, \theta, \phi)$ in which $\Omega = \omega$ and in a neighborhood of $\mathscr{I}^+$ the conformal (unphysical) metric is of the form $\tilde{g}^{(3)}_{\mu\nu}$ for $-\infty < u < \infty$, $0 \leqslant \omega < \omega_0$ and θ, ϕ as usual.

Proof: Let $\mathscr{I}^+$ be isometric to $\mathscr{I}^+_M$ and $\Omega_{;\mu\nu} \hat{=} 0$. Then a coordinate system (u,ω,θ,ϕ) exists (Sec. 2) with $\omega = \Omega$ in which on a neighborhood of $\mathscr{I}^+$ the metric is of the form $\tilde{g}^{(1)}_{\mu\nu}$. In this coordinate system the equation $\Omega_{;\mu\nu} \hat{=} 0$ gives $\beta_{00} = \beta_{01} - \frac{1}{2}\alpha_{11,0} = \beta_{02} - \alpha_{12,0} = \beta_{03} - \alpha_{13,0} = 0$. From the transformations considered in Sec. 4 we choose $u = u' + u_1\omega', \omega = \omega', \theta = \theta' + \theta_1\omega', \phi = \phi' + \phi_1\omega'$. Each one of $\alpha_{11}, \alpha_{12}, \alpha_{13}, \beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}$ is a function of u, θ, ϕ. Expanding in powers of ω' we have

$$\alpha_{12}(u,\theta,\phi) = \alpha^0_{12} + O_1,$$

$$\alpha_{12,0} \equiv \frac{\partial \alpha_{12}(u,\theta,\phi)}{\partial u} = \frac{\partial \alpha^0_{12}}{\partial u'} + O_1, \quad \text{etc.},$$

where $\alpha^0_{12} \equiv \alpha_{12}(u',\theta',\phi')$, etc. Hence we have

$$\beta^0_{00} = \beta^0_{01} - \frac{1}{2}\frac{\partial \alpha^0_{11}}{\partial u'} = \beta^0_{02} - \frac{\partial \alpha^0_{12}}{\partial u'}$$

$$= \beta^0_{03} - \frac{\partial \alpha^0_{13}}{\partial u'} = 0.$$

Because of these equations the above transformation with

$$u_1 = \tfrac{1}{2}[\alpha^0_{11} + (\alpha^0_{12})^2 + \sin^{-2}\theta\,(\alpha^0_{13})^2],$$

$$\theta_1 = \alpha^0_{12}, \quad \phi_1 = \sin^{-2}\theta'\alpha^0_{13}$$

reduces $\tilde{g}^{(1)}_{\mu\nu}$ to $\tilde{g}^{(3)}_{\mu\nu}$. This is proved easily by direct calculation of $\tilde{g}'^{(1)}_{\mu\nu}$. The converse is obvious because the hypersurface $\mathscr{I}^+$ $(\omega = 0)$ has induced metric $\tilde{\gamma}_{ij}$ and $\Omega_{;\mu\nu} \hat{=} 0$ is satisfied automatically (Sec. 3).

Theorem 2: Let $(\mathscr{M}, g)$ $(\widetilde{\mathscr{M}}, \tilde{g})$, $\mathscr{I}^+$, Ω be as Theorem 1. Then $\mathscr{I}^+$ is isometric to $\mathscr{T}^1_M$, $\Omega_{;\mu\nu} \hat{=} 0$ and $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} = -1$ if and only if there is a coordinate system $(u, \omega, \theta, \phi)$ in which $\Omega = \omega$ and in a neighborhood of $\mathscr{I}^+$ the conformal metric is of the form $\tilde{g}^{(4)}_{\mu\nu}$ for $-\infty < u < \infty$, $0 \leqslant \omega < \omega_0$, and θ, ϕ as usual.

Proof: Let $\mathscr{I}^+$ be isometric to $\mathscr{I}^+_M$, $\Omega_{;\mu\nu} \hat{=} 0$ and $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$. Then (Theorem 1) there is a coordinate system $(u, \omega, \theta, \phi)$ in which $\Omega = \omega$, and the metric is of the form $\tilde{g}^{(3)}_{\mu\nu}$. In this coordinate system the condition $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} = -1$ gives [Sec. 3, condition (iv)] $\gamma_{00} = 1$. Hence the metric is of the form $\tilde{g}^{(4)}_{\mu\nu}$. The reverse is again obvious.

From Theorems 1 and 2 we conclude that an appropriate definition of asymptotic flatness should include a demand for the conformal metric to behave as $\tilde{g}^{(3)}_{\mu\nu}$ or, better, $\tilde{g}^{(4)}_{\mu\nu}$. Unfortunately it seems that there is no similar theorem for $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$. However, if $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$, then we can still give a general *explicit* expression of the metric which automatically satisfies the conditions $\Omega_{;\mu\nu} \hat{=} 0$, $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$, and $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$. Furthermore, although $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ is not equivalent to a condition of the form $R_\mu{}^\nu = O(\Omega^n)$, we can find conditions for the physical Ricci tensor which are equivalent to $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$. From the structure of Eqs. (A15)–(A23) we observe that in order to get $A = B = C = E = F = 0$ without getting $D = G = 0$ we have to somehow substract the derivatives of $R_1{}^0$ from the other components of $R_\mu{}^\nu$. Testing several tensor expressions we find the following theorems:

Theorem 3: If a space–time satisfies conditions A and B of Sec. 2 and $\Omega_{;\mu\nu} \hat{=} 0$, then the condition $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ is equivalent to the condition $S_{\lambda\mu}{}^\nu = O(\Omega^4)$, where

$$S_{\lambda\mu}{}^\nu = \Omega_{;[\lambda}R_{\mu]}{}^\nu + \Omega R_{[\lambda}{}^\nu{}_{;\mu]}. \tag{58}$$

Proof: Since $\Omega_{;\mu\nu} \hat{=} 0$, there is a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(3)}_{\mu\nu}$. In that system the components of $S_{\lambda\mu}{}^\nu$ are given by Eqs. (A24)–(A34). Each of the conditions $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ and $S_{\lambda\mu}{}^\nu = O(\Omega^4)$ is equivalent (in this coordinate system) to the set $A = B = C = E = F = 0$. Hence these two tensor conditions are equivalent.

Therorem 4: If a space–time satisfies conditions A and B of Sec. 2, then the conditions $\Omega_{;\mu\nu} \hat{=} 0$ and $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ are equivalent to the conditions $R_\mu{}^\nu = O(\Omega^2)$ and $S_{\lambda\mu}{}^\nu = O(\Omega^4)$.

Proof: This is a direct consequence of the equivalence of the condition $\Omega_{;\mu\nu} \hat{=} 0$ to the condition $R_\mu{}^\nu \hat{=} O(\Omega^2)$ and the previous theorem.

As a result of the established properties (Secs. 3–5) we propose the following definition:

Definition: A space–time $(\mathscr{M}, \mathbf{g})$ is *asymptotically Minkowskian at future null infinity* if it is weakly asympotically simple (let Ω be the conformal factor, $\mathscr{I}$ be the hypersurface $\Omega = 0$, and $\tilde{\mathbf{g}}$ the conformal metric) and satisfies one of the following three equivalent sets of conditions:

(a) There is a subset $\mathscr{I}^+$ of $\mathscr{I}$ which is isometric to $\mathscr{I}^+_M$ and on $\mathscr{I}^+$ $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$, $\Omega_{;\mu\nu} \hat{=} 0$, $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$.

(b) There is a subset $\mathscr{I}^+$ of $\mathscr{I}$ which is isometric to $\mathscr{I}_M^+$, and on $\mathscr{I}^+$ $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$, $R_\mu{}^\nu = O(\Omega^2)$, $S_{\lambda\mu}{}^\nu = O(\Omega^4)$.

(c) There is a coordinate system $(u, \omega, \theta, \phi)$ in which on a neighborhood of a subset $\mathscr{I}^+$ of $\mathscr{I}$ we have $\Omega = \omega$ and the unphysical metric is of the form

unphysical metric is of the form

$$\tilde{g}_{\mu\nu} = \begin{bmatrix} \omega^2 + O_3 & -1 + O_2 & [\sin^{-2}\theta V_{,3} + \tfrac{1}{2}(U-\Phi)_{,2} + \cot\theta U]\omega^2 + O_3 & [V_{,2} + \cot\theta V - \tfrac{1}{2}(\Phi+U)_{,3}]\omega^2 + O_3 \\ \cdots & O_2 & O_2 & O_2 \\ \cdots & \cdots & -1 + (\Phi+U)\omega + (W - \tfrac{1}{2}\Phi U)\omega^2 + O_3 & 2V\omega - \Phi V\omega^2 + O_3 \\ \cdots & \cdots & \cdots & -\sin^2\theta + \sin^2\theta(\Phi - U)\omega + \sin^2\theta(W + \tfrac{1}{2}\Phi V)\omega^2 + O_3 \end{bmatrix}, \tag{59}$$

where $-\infty < u < \infty$, $0 \leqslant \omega < \omega_0$, θ, ϕ as usual and U, V, W are functions of u, θ, ϕ, while Φ is a function of θ, ϕ only.

Obviously in this definition the condition $\Omega_{;\mu\nu} \hat{=} 0$ can be replaced by the equivalent condition (vi) or (vii) of Sec. 3. Furthermore, since the condition $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$ has been included in the definition, the requirement that a subset $\mathscr{I}^+$ of $\mathscr{I}$ isometric to $\mathscr{I}_M^+$ exists can be weakened. If S^2 is the unit two-dimensional sphere with the usual metric imposed on it and R^1 the real line, then it is enough to assume that $\mathscr{I} \supset R^1 \times S^2$. Then there is a coordinate system (u, θ, ϕ) on $\mathscr{I}^+ \equiv R^1 \times S^2$ in which the metric $\tilde{\gamma}_{ij}$ is of the form given by Eq. (2) with perhaps $\tilde{\gamma}_{00} \neq 0$. But $\Omega_{;\mu}\Omega^{;\mu} \hat{=} 0$ from which $\tilde{\gamma}_{00} = 0$. Thus $\mathscr{I}^+$ is isometric to $\mathscr{I}_M^+$. Finally, it should be emphasized that the fact the metric can be written explicitly as in Eq. (59) is due to the tensor conditions of the definition and the property that Eqs. (26) can be solved explicitly with respect to the $\gamma_{\mu\nu}$'s they contain.

6. REMARKS AND CONCLUSIONS

Since many properties of the space–time we want to call asymptotically flat are consequences of $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ we have included this condition into the definition of asymptotic flatness.[19] Thus we propose that we distinguish four classes of space–times. In increasing order of specialization these classes are:

1. Weakly asymptotically simple space–times denoted by WASS.

2. Almost asymptotically flat space–times (AAFS) which also satisfy Condition B of Sec. 2 and $\Omega_{;\mu\nu} = 0$,

$\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$.

3. Asymptotically flat space–times (AFS), that is which also satisfy the condition $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$.

4. Asymptotically flat and empty space–times (AFES), that is AFS which also satisfy the condition $R_\mu{}^\nu = O(\Omega^4)$.

The definition given in the previous section refers to the third class. The corresponding physical metric in coordinates (u, r, θ, ϕ) can be compared to metrics used in previous works. The metrics used by Bondi *et al.*[1] and Sachs[5] have $\alpha_{11} = \alpha_{12} = \alpha_{13} = \beta_{00} = \beta_{01} = \beta_{02} = \beta_{03} = 0$, $\gamma_{00} = 1$ and hence satisfy $\Omega_{;\mu\nu} \hat{=} 0$ and $\Omega^{-2}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -1$ as a result of their construction. The metric used by Persides[6,8] has α_{11}, $\alpha_{12}, \alpha_{13}, \beta_{01}$ different from zero. Hence in general it does not represent asymptotically flat space–time. However, if we assume the Einstein field equations, then the coefficients satisfy Eqs. (19) and the space–time becomes asymptotically flat.

Properties of asymptotically flat space–times can be proved easily in a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(4)}_{\mu\nu}$. Thus if

$$L_{\mu\nu} = R_{\mu\nu} + \kappa R g_{\mu\nu}, \quad \widetilde{L}_{\mu\nu} = \widetilde{R}_{\mu\nu} + \kappa \tilde{R}\tilde{g}_{\mu\nu}, \tag{60}$$

where κ is an arbitrary constant, then straightforward calculations on $\mathscr{I}^+$ give

$$L_{\mu\nu}\Omega^{;\nu} \hat{=} 0, \quad \widetilde{L}_{\mu\nu}\Omega^{;\nu} \hat{=} -\Omega_{;\rho}, \tag{61}$$

$$\widetilde{L}_{\lambda\mu;\nu}\Omega^{;\mu}\Omega^{;\nu} \hat{=} 0, \quad \Omega_{;\mu\nu\rho}\Omega^{;\mu} \hat{=} -\Omega_{;\nu}\Omega_{;\rho}. \tag{62}$$

Properties of almost asymptotically flat space–times can be established in a similar way. For example, we can prove that for an AAFS energy can be defined iff $\Omega^{-1}\Omega_{;\mu}\Omega^{;\mu} \hat{=} -2$ or $\widetilde{R} \hat{=} 0$. The proof is simple in a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(4)}_{\mu\nu}$, since there Eq. (14) or Eq. (34) imply Eq. (15) which is the necessary and sufficient condition for the convergence of the energy integral (then and only then the integrand[6] starts with an $\omega^2 = r^{-2}$ term).

Throughout this work we studied asymptotic flatness at future null infinity. Past null infinity can be studied in a similar way. Also the global requirement that $\mathscr{I}^+$ be isometric to $\mathscr{I}_M^+$ can be relaxed to incorporate space–times which have a part of $\mathscr{I}^1$ removed. Thus we can have partially AFS which satisfy the requirements of the AF space–time on some three-dimensional subset $\mathscr{I}_p^+$ of $\mathscr{I}$ such that $\mathscr{I}_p^+$ is isometric to a subset of $\mathscr{I}_M^+$.

ACKNOWLEDGMENTS

I would like very much to thank Dr. J. Stewart and Dr.

B. Xanthopoulos for commenting on a preliminary version of this paper. I would also like to thank the members and staff of the Institute of Astronomy at the University of Cambridge (where most of this work was done) for their hospitality and support.

APPENDIX

Let $g_{\mu\nu}$ be the physical metric and $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ the unphysical metric as given by Eq. (6). We define

$$\Phi = \tfrac{1}{2}\beta_{22} + \frac{1}{2\sin^2\theta}\beta_{33}, \tag{A1}$$

$$U = \tfrac{1}{2}\beta_{22} - \frac{1}{2\sin^2\theta}\beta_{33}, \tag{A2}$$

$$V = \tfrac{1}{2}\beta_{23}, \tag{A3}$$

$$W = \tfrac{1}{2}\gamma_{22} + \frac{1}{2\sin^2\theta}\gamma_{33}, \tag{A4}$$

$$A = \gamma_{00} - 1 + \Phi_{,0}, \tag{A5}$$

$$B = \gamma_{02} - \frac{1}{\sin^2\theta}V_{,3} + \tfrac{1}{2}(\Phi - U)_{,2} - \cot\theta U - \tfrac{1}{2}\beta_{12,0}, \tag{A6}$$

$$C = \gamma_{03} - V_{,2} - \cot\theta V + \tfrac{1}{2}(\Phi + U)_{,3} - \tfrac{1}{2}\beta_{13,0}, \tag{A7}$$

$$D = 4\gamma_{01} + 2W + \tfrac{1}{2}(\Phi^2 + U^2) + \frac{2}{\sin^2\theta}V^2 - \beta_{12,2} - \frac{\beta_{13,3}}{\sin^2\theta} - \beta_{12}\cot\theta - \beta_{11,0}, \tag{A8}$$

$$E = \tfrac{1}{2}\gamma_{22} - \frac{1}{2\sin^2\theta}\gamma_{33} + \tfrac{1}{2}\Phi U - \tfrac{1}{2}\beta_{12,2} + \frac{1}{2\sin^2\theta}\beta_{13,3} + \tfrac{1}{2}\cot\theta\beta_{12}, \tag{A9}$$

$$F = \gamma_{23} + \Phi V - \tfrac{1}{2}\beta_{12,3} - \tfrac{1}{2}\beta_{13,2} + \cot\theta\beta_{13}, \tag{A10}$$

$$G = 6\delta_{01} + 3\delta_{22} + \frac{3}{\sin^2\theta}\delta_{33} + (\text{terms with } \beta_{\mu\nu} \text{ and } \gamma_{\mu\nu} \text{ only}). \tag{A11}$$

If there is a coordinate system in which $\tilde{g}_{\mu\nu} = \tilde{g}^{(3)}_{\mu\nu}$, then straightforward calculations in this coordinate system give the (unphysical) Weyl tensor

$$3\widetilde{C}_{0101} \overset{\text{cf25}\xi}{\widehat{=}} 6\widetilde{C}_{0212} \widehat{=} 6\sin^{-2}\theta\widetilde{C}_{0313} \overset{\text{cf25}\xi}{\widehat{=}} -2\sin^{-2}\theta\widetilde{C}_{2323} \widehat{=} A, \tag{A12}$$

$$2\widetilde{C}_{0112} \widehat{=} -2\sin^{-2}\theta\widetilde{C}_{1323} \widehat{=} -B, \quad 2\widetilde{C}_{0113} \widehat{=} 2\widetilde{C}_{1223} \widehat{=} -C, \tag{A13}$$

$$\widetilde{C}_{1212} \widehat{=} -\sin^{-2}\theta\widetilde{C}_{1313} \widehat{=} E, \quad \widetilde{C}_{1213} \widehat{=} F, \tag{A14}$$

with zero the remaining components. In the same coordinate system $g_{\mu\nu} = g^{(3)}_{\mu\nu} = \Omega^{-2}\tilde{g}^{(3)}_{\mu\nu}$ and the (physical) Ricci tensor can be written (with $\Omega = \omega$)

$$R_\mu{}^\nu = C_\mu{}^\nu\omega^2 + D_\mu{}^\nu\omega^3 + O_4 \tag{A15}$$

with

$$C_2{}^2 = C_3{}^3 = -A, \quad C_2{}^0 = C_1{}^2 = B, \quad C_3{}^0 = \sin^2\theta C_1{}^3 = C, \quad C_1{}^0 = D, \tag{A16}$$

$$D_0{}^0 = D_1{}^1 = \tfrac{1}{2}D_{,0}, \quad D_0{}^1 = A_{,0}, \quad D_0{}^2 = D_2{}^1 = B_{,0}, \tag{A17}$$

$$D_0{}^3\sin^2\theta = D_3{}^1 = C_{,0}, \quad D_1{}^0 = G, \quad D_3{}^2 = \sin^2\theta D_2{}^3 = F_{,0}, \tag{A18}$$

$$D_2{}^2 - E_{,0} = D_3{}^3 + E_{,0} = -\Phi A + B_{,2} + \cot\theta B + \sin^{-2}\theta C_{,3}, \tag{A19}$$

$$D_2{}^0 = \Phi B + \tfrac{1}{2}D_{,2} - E_{,2} - 2\cot\theta E - \sin^{-2}\theta F_{,3}, \tag{A20}$$

$$D_3{}^0 = \Phi C + \tfrac{1}{2}D_{,3} + E_{,3} - F_{,2} - \cot\theta F, \tag{A21}$$

$$D_1{}^2 = (2\Phi + U)B + \beta_{12}A + \tfrac{1}{2}D_{,2} + \sin^{-2}\theta\beta_{23}C - E_{,2} - 2\cot\theta E - \sin^{-2}F_{,3}, \tag{A22}$$

$$D_1{}^3\sin^2\theta = (2\Phi - U)C + \beta_{13}A + \tfrac{1}{2}D_{,3} + \beta_{23}B + E_{,3} - F_{,2} - \cot\theta F, \tag{A23}$$

with the remaining $C_\mu{}^\nu$ equal to zero. Finally for $S_{\lambda\mu}{}^\nu$ we have

$$S_{\mu\nu}{}^\nu = \mathscr{C}_{\lambda\mu}{}^\nu\omega^2 + \mathscr{D}_{\lambda\mu}{}^\nu\omega^3 + O_4 \tag{A24}$$

with

$$\mathscr{C}_{12}{}^0 = -\tfrac{1}{2}B, \quad \mathscr{C}_{13}{}^0 = -\tfrac{1}{2}C, \quad \mathscr{C}_{12}{}^2 = \mathscr{C}_{13}{}^3 = \tfrac{1}{2}A, \tag{A25}$$

$$\mathscr{D}_{02}{}^2 = -\mathscr{D}_{01}{}^2 = \tfrac{1}{2}\mathscr{D}_{12}{}^1 = -\tfrac{1}{2}B_{,0}, \quad \mathscr{D}_{03}{}^0 = -\sin^2\theta\mathscr{D}_{01}{}^3 = \tfrac{1}{2}\mathscr{D}_{13}{}^1 = -\tfrac{1}{2}C_{,0}, \tag{A26}$$

$$\mathscr{D}_{01}{}^1 = \mathscr{D}_{02}{}^2 = \mathscr{D}_{03}{}^3 = \tfrac{1}{2}A_{,0}, \quad \mathscr{D}_{23}{}^2 = -\tfrac{1}{2}A_{,3}, \quad \mathscr{D}_{23}{}^3 = \tfrac{1}{2}A_{,2}, \tag{A27}$$

$$\mathscr{D}_{12}{}^2 = \tfrac{1}{4}(3\Phi - U)A - B_{,2} - \cot\theta B - \sin^{-2}\theta C_{,3} - E_{,0}, \tag{A28}$$

$$\mathscr{D}_{13}{}^3 = \tfrac{1}{4}(3\Phi + U)A - B_{,2} - \tfrac{1}{2}\cot\theta B - \tfrac{1}{2}\sin^{-2}\theta C_{,3} + E_{,0}, \tag{A29}$$

$$\mathscr{D}_{13}{}^2 = -\tfrac{1}{2}VA + \tfrac{1}{2}B_{,3} - \tfrac{1}{2}\cot\theta C - F_{,0}, \tag{A30}$$

$$\mathscr{D}_{12}{}^3 = -\tfrac{1}{2}VA + \tfrac{1}{2}\sin^{-2}\theta C_{,2} - \tfrac{1}{2}\cot\theta C - \tfrac{1}{4}\sin^{-2}\theta F_{,0}, \tag{A31}$$

$$\mathscr{D}_{12}{}^0 = -\Phi B + \tfrac{1}{2}\sin^{-2}\theta VC + E_{,2} + 2\cot\theta E + \sin^{-2}\theta F_{,3}, \tag{A32}$$

$$\mathscr{D}_{13}{}^0 = -\Phi C + \tfrac{1}{2}VB - E_{,3} + F_{,2} + \cot\theta F, \tag{A33}$$

$$\mathscr{D}_{23}{}^0 = \tfrac{1}{2}VA + \tfrac{1}{2}B_{,3} - \tfrac{1}{2}C_{,2}, \tag{A34}$$

with the remaining $\mathscr{C}_{\lambda\mu}{}^\nu$ and $\mathscr{D}_{\lambda\mu}{}^\nu$ equal to zero.

[1]H. Bondi, M.G.J. van der Burg, and A.W.K. Metzner, Proc. Roy. Soc. A **269**, 21 (1962).
[2]R. Penrose, Proc. Roy. Soc. A **284**, 159 (1965).
[3]R. Penrose, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).
[4]R. Penrose, in *Battelle Rencontres, Lectures in Mathematical Physics*, edited by C. DeWitt and J.A. Wheeler (Benjamin, New York, 1968).
[5]R.K. Sachs, Proc. R. Soc. A **270**, 103 (1962).
[6]S. Persides, Proc. R. Soc. A **320**, 349 (1970).
[7]M.G.J. van der Burg, Proc. Roy. Soc. A **310**, 221 (1969).
[8]S. Persides, Prog. Theor. Phys. **58**, 829 (1977).
[9]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).
[10]If $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$, then $\widetilde{K}_{\lambda\mu\rho\nu} = \Omega^{-1}\widetilde{C}_{\lambda\mu\rho\nu}$ is a smooth tensor field on $\mathscr{I}$ and contains information about radiation, etc. Even properties of the interior of the space–time depend on the existence of $\widetilde{K}_{\lambda\mu\rho\nu}$. See, e.g., A. Ashtekar and B. Xanthopoulos, J. Math. Phys. **19**, 2216 (1978); also Ref. 14.
[11]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U. P., Cambridge, 1973).
[12]R. Geroch, and G.T. Horowitz, Phys. Rev. Lett, **40**, 203 (1978). We adopt should include conditions on the asymptotic regime only.
[13]The usual definition of weak asymptotic simplicity is based on the existence of an asymptotically simple space–time $(\mathscr{M}', \mathbf{g}')$ which has a neighborhood of $\mathscr{I}'$ isometric to a neighborhood of $\mathscr{I}$. In general, however, it is not obvious that $\mathscr{M}'$ exists always and a definition of weak asymptotic simplicity which does not depend on $\mathscr{M}'$ seems to be preferable.
[14]R. Geroch, in *Asymptotic Structure of Space–Time*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1977). This recent review article contains most of the background used in this paper.
[15]The condition (Ref. 12) that $\Omega^{,\mu}$ be a complete vector field on $\mathscr{I}^+$ follows from the isometry of $\mathscr{I}^+$ and $\mathscr{I}^+_M$ and the fact that $\Omega^{,\mu}$ is complete on $\mathscr{I}^+_M$.
[16]L.A. Tamburino and J.H. Winicour, Phys. Rev. **150**, 1039 (1966).
[17]The equation $G = 0$ determines uniquely the combination $2\delta_{01} + \delta_{22} + \sin^{-2}\theta\delta_{33}$ in terms of $\beta_{\mu\nu}$ and $\gamma_{\mu\nu}$.
[18]R. Sachs, Phys. Rev. **128**, 2851 (1962).
[19]Geroch and Horowitz (Ref. 12) do not include $\Omega^{-2}\Omega_{,\mu}\Omega^{,\mu} \hat{=} -1$ and $\widetilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ in the definition. Essentially they define as asymptotically flat space–time what we call "almost asymptotically flat."

A necessary and sufficient condition for York data to specify an asymptotically flat spacetime

Murray Cantor[a)]
Institute for Advanced Study, Princeton, New Jersey 08540
and Department of Mathematics, University of Texas, Austin, Texas 78712

(Received 19 January 1979)

This paper studies the conditions under which Cauchy data $(\bar{g},\bar{\pi},\bar{\nu},\bar{T})$ for an asymptotically flat spacetime are determined by the freely specifiable York data (g, σ, ν, T) $(\tau = 0)$, where $\mathrm{tr}_g\sigma = 0$, $\mathrm{div}_g\sigma = 8\pi\nu$. It is shown that the space of such σ's is infinite dimensional. Furthermore,it is shown that (g, σ, ν, T) determine conformally equivalent Cauchy data if and only if g is conformally equivalent to an asymptotically flat metric with nonnegative scalar curvature.

INTRODUCTION

In the Cauchy formulation for constructing a spacetime satisfying the Einstein field equations, $G_{\mu\nu} = 8\pi T_{\mu\nu}$, and containing a spacelike hypersurface M, one specifies the following initial data on M (see Choquet-Bruhat and York, [1] York, [2] or Fischer and Marsden [3] for various general treatments of the Cauchy problem):

$\bar{g}$–a Riemannian metric on M,
$\bar{\pi}$–a symmetric covariant 2-tensor on M,
$\bar{\nu}$–a vector field on M,
$\bar{T}$–a positive function on M.

(In some formulations $\bar{\pi}$ is given as a tensor density; we will use tensors throughout the paper. Also, unless explicitly written otherwise, we will treat all tensors in the form with all indices lowered.) $\bar{\pi}$ is the momentum conjugate to the $\bar{g}$'s and is related to the second fundamental form k of the embedding of M in the spacetime by the formula $\bar{\pi} = (\mathrm{tr}_{\bar{g}}k)\bar{g} - k$. $\bar{\nu}$ is the current density of $T_{\mu\nu}$ on M and $\bar{T}$ is the energy density of $T_{\mu\nu}$ on M (hence $\bar{T}\geqslant 0$).

The initial data $(\bar{g},\bar{\pi},\bar{\nu},\bar{T})$ are not freely specifiable but must satisfy the constraint equations:

$$\mathrm{div}_{\bar{g}}\bar{\pi} = -8\pi\bar{\nu}, \quad R(\bar{g}) + \tfrac{1}{2}(\mathrm{tr}_{\bar{g}}\bar{\pi})^2 - \bar{\pi}\cdot\bar{\pi} = 16\pi\bar{T}. \qquad \text{(C)}$$

Our notation using the summation convention is $\mathrm{div}_g k = k^{ab}{}_{|b}$, $k\cdot k = k^{ab}k_{ab}$, $\mathrm{tr}_g k = g^{ab}k_{ab}$, and $R(g)$ is the scalar curvature of g. Differentiation is taken with respect to g.

Using conformal methods, York determined the freely specifiable "pieces" of the initial data (York, [2] York, [4] O'Murchadha and York [5] and Choquet-Bruhat and York[1]). These are

g–a conformal metric on M,
σ–a symmetric 2-tensor satisfying $\mathrm{tr}_g\sigma = 0$, $\mathrm{div}_g\sigma = 8\pi\nu$,
ν–a vector field on M,
T–a positive function on M,
τ–a function on M.

τ is the mean curvature of the embedding of M in the constructed spacetime. (g,σ,ν,T,τ) is called York data on M.

If it is assumed τ is constant on M, one can determine an initial data set $(\bar{g},\bar{\pi},\bar{\nu},\bar{T})$ from York data (g,σ,ν,T,τ) by solving

$$8\Delta_g\varphi - R(g)\varphi + M\varphi^{-7} + 16\pi T\varphi^{-3} + \tfrac{2}{3}\tau^2\varphi^5 = 0, \qquad (1)$$

where $\varphi > 0$, $\Delta_g\varphi = g^{ab}\varphi_{|ab}$, and $M = \sigma\cdot\sigma \geqslant 0$. Then $\bar{g} = \varphi^4 g$, $k = \varphi^{-2}\sigma + \frac{1}{3}\varphi^4\tau g$, $\bar{\pi} = (\mathrm{tr}_{\bar{g}}k)\bar{g} - k$, $\bar{\nu} = \varphi^{-10}\nu$, and $\bar{T} = \varphi^{-8}T$. Note that g is given only up to conformal equivalence. Because of this, the initial data $(\bar{g},\bar{\pi},\bar{\nu},\bar{T})$ is uniquely determined by the York data in the following sense: If we start with $g' = \alpha^4 g$, and set $\sigma' = \alpha^{-10}\sigma$, $\nu' = \alpha^{-10}\nu$, $T' = \alpha^{-8}T$, and $\tau' = \tau$, then (g,σ,ν,T,τ) and $(g',\sigma',\nu',T',\tau')$ determine the same initial data (York [2,4]).

In the asymptotically flat case with $M = \mathbb{R}^3$, it is usually assumed that $\tau = 0$. This allows M to be spacelike at infinity. (For a discussion of asymptotically flat spacetimes allowing such spacelike hypersurfaces, see Cantor *et al.* [6] and Tipler and Marsden. [7]) Thus, the York data will from now on be given as a 4-tuple (g,π,ν,T). Also, certain growth conditions are assumed as $|x|\to\infty$ (see Regge-Tietelboim [8]):

$$g_{ij} - \delta_{ij} \sim |x|^{-1},$$
$$\nu, T, R \sim |x|^{-3},$$
$$\sigma \sim |x|^{-2}.$$

Also, we require that the initial data $(\bar{g},\bar{\pi},\bar{\nu},\bar{T})$ preserve these conditions. We accomplish this by finding a solution to (1) with the property $\varphi - 1 \sim |x|^{-1}$. Using weighted Sobolev spaces we will make these conditions more precise below.

In order to carry out York's program we need to solve two problems:

1: Does there exist nontrivial σ for a given metric g? 2: Can we use Eq. (1) to find to conformal factor? Using a result found in Cantor, [9] it is easy to show the space of σ's is in fact infinite dimensional (see below) and so problem 1 presents no difficulty.

Problem 2 is more difficult. Equation (1) is very nonlinear. However, we do have the conformal invariance described above at our disposal. In solving problem 2 there is

a)This research was partially supported by a National Science Foundation Research Grant.

 0022-2488/79/081741-04$01.00

an immediate necessary condition. If we start with a metric g and construct initial data containing $\bar{g} = \varphi^4 g$ we find from Eqs. (C) that when $\tau = 0$ then $\bar{R} = \bar{\pi}\cdot\bar{\pi} + 16\pi T \geqslant 0$. Thus it is necessary that our initial g be conformally equivalent to one with the appropriate growth at infinity and possessing nonnegative scalar curvature. We show below that the above necessary condition is sufficient. This represents a substantial improvement on previously known results (see Cantor, [9] Chaljub and Choquet-Bruhat [10]). Previously, certain inequalitites between the R, M, and T coefficients were needed to show (1) may be solved. This paper is the first giving both necessary and sufficient conditions for the York variables to specify initial data.

One question that is left unresolved is the following: Is any metric g_{ij} on $\mathbb{R}^3$ with $g_{ij} - \delta_{ij} \sim |x|^{-1}$ as $|x| \to \infty$ conformally equivalent to one with the same asymptotic conditions and nonnegative scalar curvature? If this is resolved in the positive then we would know that in the asymptotically flat case every set of York variables do specify initial data.

1. WEIGHTED SOBOLEV SPACES

In order to make precise the asymptotic conditions on the various pieces of the data and also to establish certain necessary technical lemmas, we introduce the following weighted Sobolev spaces:

Definition 1.1: Let $p \geqslant 1$, $s \in \mathbb{N}$ and $\delta \in \mathbb{R}$, and $\sigma(x) = (1 + |x|^2)^{1/2}$. For $f: \mathbb{R}^n \to \mathbb{R}^m$ set $|f|_{p,s,\delta} = \Sigma_{|\alpha| \leqslant s} |\sigma^{\delta + |\alpha|} D^\alpha f|_p$ where $|\ |_p$ is the standard L^p norm. $M^p_{s,\delta}(\mathbb{R}^n, \mathbb{R}^m)$ is the completion of $C^\infty_0(\mathbb{R}^n, \mathbb{R}^m)$ with respect to $|\ |_{p,x,\delta}$.

We can specify the asymptotic behavior of a function by assuming it belongs to the appropriate $M^p_{s,\delta}$ space (see below). But first we state some lemmas which we shall need later.

Lemma 1.2: *Let $p > 1$, $s > n/p$, $\delta \geqslant 0$ and $0 \leqslant l \leqslant s$. Then pointwise multiplication induces a continuous (and hence smooth) map*:

$$M^p_{s,\delta} \times M^p_{s-l,\delta+l} \to M^p_{s-l,\delta+l}.$$

For a proof see Cantor, [11] Proposition 1.1.

An immediate consequence is the following:

Lemma 1.3: *Let $M^p_{s,\delta}(1) = \{ g: \mathbb{R}^3 \to \mathbb{R}, g - 1 \in M^p_{s,\delta} \}$ be given the topology such that the map $g \to g + 1$ from $M^p_{s,\delta}$ is continuous. Then for $p > 1$, $s > 3/p$, $\delta \geqslant 0$, $0 \leqslant l \leqslant s$, multiplication induces smooth maps*:

$$\begin{aligned} &M^p_{s,\delta}(1) \times M^p_{s-l,\delta+l}(1) \to M^p_{s-l,\delta+l}(1), \\ &M^p_{s,\delta} \times M^p_{s-l,\delta+l}(1) \to M^p_{s-l,\delta+l}. \end{aligned} \tag{2}$$

Definition 1.4: Let $S^p_{s,\delta} = \{$symmetric covariant 2-tensors in $M^p_{s,\delta}\}$ and $R^p_{s,\delta} = \{$Reimannian metrices g_{ij} on $\mathbb{R}^3$ such that $g_{ij} - \delta_{ij} \in S^p_{s,\delta}\}$.

We will need the following theorems concerning elliptic operators.

Theorem 1.5: *Let $p > 3$, $s \geqslant 3$, $0 \leqslant \delta < -2 + 3(p-1)/p$, $g \in R^p_{s,\delta}$ and $f \in M^p_{s-2,\delta+2}$ with $f \geqslant 0$. Then if $A(\varphi) = \Delta_g \varphi - f\cdot\varphi$ and $2 \leqslant l \leqslant s$ we have $A: M^p_{l,\delta} \to M^p_{-2,\delta+2}$ is an isomorphism.*

Proof: This is a simple extension of Theorem 1.4 in Cantor. [12]

Theorem 1.6 (The maximum principle): *If φ is a C^2 function and $\Delta_g \varphi \geqslant 0$* (resp. $\leqslant 0$), *then φ cannot have a maximum (resp. minimum) in any open region in $\mathbb{R}^3$, unless φ is constant.*

Proof: This is a standard result. See for example Protter and Weinberger. [13]

Theorem 1.7: *In a region G let a second order elliptic linear differential equation have two C^1 solutions $u(x)$ and $v(x)$. Suppose they satisfy the conditions*:

(i) *throughout G, $u(x) \geqslant v(x)$;*

(ii) *at a point $x_0 \in G$ such that there exits a closed sphere $S \subset G$ containing x_0 in its interior or on the boundary, $u(x_0) = v(x_0)$ and $\nabla u(x_0) = \nabla v(x_0)$.*

Then $u(x) = v(x)$ throughout G.

Proof: This is a special case of Theorem A in Aleksandrov. [14]

2. INITIAL DATA SETS

In light of Theorem 1.5 we will make the following choices which will be used throughout the paper:

$$\begin{aligned} &p > 3, \\ &s \geqslant 3, \\ &0 \leqslant \delta < -2 + 3(p-1)/p. \end{aligned}$$

With these choices we may specify the asymptotic conditions mentioned in the Introduction.

Definition 2.1: An asymptotically flat York data set is a 4-tuple (g, σ, ν, T) where $g \in R^p_{s,\delta}$ (recall Definition 1.4), $\sigma \in S^p_{s-1,\delta+1}$ with $\mathrm{tr}_g \sigma = 0$ and $\mathrm{div}_g \sigma = 8\pi\nu$, $\nu \in M^p_{s-2,\delta+2}(\mathbb{R}^3, \mathbb{R}^3)$ and $T \in M^p_{s-2,\delta+2}(\mathbb{R}^3, \mathbb{R})$ with $T \geqslant 0$.

To show such data exists and to solve problem 1 of the Introduction, we have the following theorem:

Theorem 2.2: *Let $g \in R^p_{s,\delta}$ and $\nu \in M^p_{s-2,\delta+2}$. Then the space of solutions to*

$$\mathrm{tr}_g \sigma = 0, \quad \mathrm{div}_g \sigma = 8\pi\nu, \tag{3}$$

in an infinite dimensional affine subspace of $S^p_{s-1,\delta+1}$.

Proof: Recall the definition of the conformal Killing operator L_g for vector fields on $\mathbb{R}^3$: $L_g X = K_g X - \frac{2}{3}(\mathrm{div}_g X) g$ where $(K_g X)_{ij} = (L_X g)_{ij} = X_{i|j} + X_{j|i}$. We will require the following two lemmas:

Lemma A: $\mathrm{div}_g \circ L_g M^p_{s,\delta} \to M^p_{s-2,\delta+2}$ *is an isomorphism.*

This is proven in Cantor. [12]

Lemma B: *The space of solutions to* (3) *with $\nu = 0$ is infinite dimensional.*

This is found as Theorem 3.7 in Cantor. [9]

solution of the form $k = L_g W$ with $W \in M^p_{s,\delta}$. Note that any such 2-tensor satisfies $\mathrm{tr}_g k = \mathrm{tr}_g(L_g W) = 0$. Now using Lemma A, write $W = (\mathrm{div}_g \circ L_g)^{-1}(\mathrm{div}_g 8\pi\nu) \in M^p_{s,\delta}$. It is clear that $k = L_g W$ satisfies (3).

Now consider $J = \{k + j: \mathrm{div}_g j = \mathrm{tr}_g j = 0, j \in S^p_{s-1,\delta+1}\}$. From Lemma B we see J is the desired infinite dimensional affine subspace of $S^p_{s-1,\delta+1}$. Q.E.D.

We now come to the main result:

Theorem 2.3: *Let (g, σ, ν, T) be an asymptotically flat*

York data set. The following statements are equivalent:

(I) *There is a* $g' \in R^{p}_{s,\delta}$ *conformally equivalent to* g *with nonnegative scalar curvature.*

(II) *There is an asymptotically flat initial data set* $(\bar{g},\bar{\pi},\bar{T},\bar{\nu})$ *which is conformally equivalent to* (g,σ,T,ν). *Moreover* $(\bar{g},\bar{\pi},\bar{T},\bar{\nu})$ *depends smoothly on* (g,σ,T,ν).

Proof [(II)→(I)]: We have that $\bar{g}$ is conformally equivalent to g and that $\bar{g},\bar{\pi},\bar{T}$ must satisfy Eq. (C) in the Introduction. Thus $R(\bar{g}) + \frac{1}{2}(\mathrm{tr}_{\bar{g}}\bar{\pi})^2 - \bar{\pi}\cdot\bar{\pi} = 16\pi\bar{T}$. Also recall $\bar{\pi} = \mathrm{tr}_{\bar{g}} k\bar{g} - k$ where $k = \varphi^{-2}\sigma$ ($\bar{g} = \varphi^4 g$). Thus $\mathrm{tr}_{\bar{g}} k = \bar{g}^{ij}k_{ij} = \varphi^{-6} g^{ij}\sigma_{ij} = 0$ and so $\mathrm{tr}_{\bar{g}}\bar{\pi} = 0$. Thus since $\bar{T} \geqslant 0$ we have $R(\bar{g}) \geqslant 0$.

[(I)→(II)]: Recall from the Introduction the desired conformal factor φ must satisfy the following problem:

$$\begin{aligned} &8\Delta_g\varphi - R(g)\varphi + M\varphi^{-7} + 16\pi T\varphi^{-3} = 0, \\ &\varphi - 1 \in M^{p}_{s,\delta}, \\ &\varphi > 0. \end{aligned} \tag{4}$$

The condition $\varphi - 1 \in M^{p}_{s,\delta}$ implies after application of Lemma 1.3 that $\bar{g} \in R^{p}_{s,\delta}$, etc. Thus the proof of the theorem results on showing (I) is sufficient for (4) to have a solution.

Step 1: *Let* g *satisfy* (I). *Then there is a* $\alpha \in M^{p}_{s,\sigma}(1)$ *such that* $g' = \alpha^4 g$ *has zero scalar curvature.*

The desired function α must satisfy the following problem (see O'Murchadhu and York [16])

$$\begin{aligned} &8\Delta_g\alpha - R(g)\alpha = 0, \\ &\alpha > 0, \\ &\alpha - 1 \in M^{p}_{s,\delta}. \end{aligned} \tag{5}$$

Using condition (I) we may assume $R(g) \geqslant 0$. Writing $\alpha = 1 + \bar{\alpha}$ we find

$$8\Delta_g\bar{\alpha} - R(g)\bar{\alpha} = \bar{R}(g).$$

It follows immediately from Theorem 1.5 that there is such a $\bar{\alpha} \in M^{p}_{s,\delta}$. We need only show $\alpha = 1 + \bar{\alpha} > 0$ on all of $\mathbb{R}^3$. Suppose there were an $x_0 \in \mathbb{R}^3$ such that $\alpha(x_0) < 0$. We can suppose x_0 is a minimum for α. In a neighborhood of x_0, we have $8\Delta_g\alpha = R(g)\alpha \leqslant 0$ and so α cannot take a minimum. Thus $\alpha \geqslant 0$. If there were an x_1 such that $\alpha(x_1) = 0$, then x_1 would be a minimum and so $\nabla\alpha(x_1) = 0$. However $\psi = 0$ is also a solution to the equation. Hence from Theorem 1.7 we have $\alpha = 0$ everywhere. However, $\alpha(x) \to 1$ as $|x| \to \infty$. Thus we have a contradiction. Hence $\alpha > 0$. This ends the proof of step 1.

Step 2: Let $M,T \in M^{p}_{s-2,\delta+2}$ and $g \in R^{p}_{s,\delta}$ with $R(g) = 0$. Then (4) has a unique solution φ. Furthermore, $\varphi \geqslant 1$.

Proof: We first show (4) has at most one solution. Note if $\varphi - 1 \in M^{p}_{s,\delta}$, then φ is C^2 and we may apply a maximum principle. Suppose φ_1 and φ_2 are both solutions to (S). Let $h = \varphi_1 - \varphi_2$. Then $h \in C^2$ and $h(x) \to 0$ as $|x| \to \infty$. Assume without loss of generality that $h > 0$ somewhere. Then there is an $x_0 \in \mathbb{R}^3$ such that $h(x_0) > 0$ and x_0 is a maximum for h. Now

$$8\Delta_g h = -M(\varphi_1^{-7} - \varphi_2^{-7}) - 16\pi T(\varphi_1^{-3} - \varphi_2^{-3})$$

and $\varphi_1(x_0) - \varphi_2(x_0) > 0$ and $\varphi_1^{-t}(x_0) - \varphi_2^{-t}(x_0) < 0$ for $t > 0$. It follows that $\Delta_g h(x) > 0$ in a neighborhood of x_0, and so x_0 cannot be a maximum. Thus $h = 0$.

To see $\varphi \geqslant 1$, note since $\mu \to 1$ as $|x| \to \infty$ and $\varphi \in C^2$ then $\Delta_g\varphi$ is continuous and hence $\varphi > 0$ everywhere. It follows that $\Delta_g\varphi \leqslant 0$ on $\mathbb{R}^3$ and so by Theorem 1.6, φ cannot take a minimum. However, if $\varphi < 1$ somewhere it must have a minimum.

To show existence we use a continuity argument. Let $S = \{t \in [0,1]$: (4) has a solution for $s \in [0,t]$ using sM and $sT\}$. We will show $S = [0,1]$. We know for $s = 0$ that $\varphi = 1$ is a solution. Hence $0 \in S$. We now show S is open in $[0,1]$.

Consider the operator

$$\Psi: M^{p}_{s,\delta}(1) \times M^{p}_{s-2,\delta+2} \times M^{p}_{s-2,\delta+2} \to M^{p}_{s-2,\delta+2}$$

(see Lemma 1.3) where $\Psi(\varphi,M,T) = \Delta_g\varphi + M\varphi^{-7} + 16\pi T\varphi^{-3}$. It follows from Lemma 1.3 and standard results on the inverse in a Banach algebra (see Dunsford and Schwartz [15]) that Ψ is C^∞ at any (φ,M,T) if $\varphi \geqslant c > 0$.

To show S is open, we use the implicit function theorem. Let $t \in S$. Then by definition for $\bar{t} \in [0,t]$ there is an $\varphi_{\bar{t}} \in M^{p}_{s,\delta}(1)$ such that $\Psi(\varphi_{\bar{t}}, \bar{t}M, \bar{t}T) = 0$. We also know $\varphi_{\bar{t}} \geqslant 1$ and so Ψ is C^∞ at $(\varphi_{\bar{t}}, tM, tT)$. To show the equation $\Psi = 0$ may be solved locally we need show $D_1\Psi(\varphi_{\bar{t}}, tM, tT): M^{p}_{2,\delta} \to M^{p}_{s-2,\delta+2}$ is an isomorphism. However,

$$D_1\Psi(\varphi_{\bar{t}}, tM, tT)(f) = \Delta_g f - [7tM\varphi_{\bar{t}}^{-8} + 48\pi tT\varphi_{\bar{t}}^{-4}]f.$$

It follows from Theorem 1.5 that this is an isomorphism. Thus (4) may be solved for slightly larger t and so S is open.

There is a maximal interval of the form $[0,t_0)$ in S. We need only show (4) is solvable at t_0. For then $[0,t_0] \subset S$ and the only open set in $[0,1]$ of this form is $[0,1]$.

For $t \in [0,t_0)$ there is a $\varphi_t \in M^{p}_{s,\delta}(1)$ such that $\Psi(\varphi_t, tM\,tT) = 0$. It follows from the implicit function argument given above that $t \to \varphi_t$ is C^∞.

Claim 1: $\{\varphi_t : t \in [0,t_0)\}$ is uniformly bounded in $M^{p}_{s,\delta}(1)$.

We write $\varphi_t = \bar{\varphi}_t + 1$ and find $\bar{\varphi}_t$ satisfying the problem:

$$\Delta_g\bar{\varphi}_t = -t[M(\bar{\varphi}_t + 1)^{-7} + T(\bar{\varphi}_t + 1)^{-3}].$$

Now $\{t[M(\bar{\varphi}_t + 1)^{-7} + T(\bar{\varphi}_t + 1)^{-3}]\}$ is clearly bounded in $M^{p}_{0,\delta+2}$. But from Theorem 1.5, $(\Delta_g)^{-1}: M^{p}_{0,\delta+2} \to M^{p}_{2,\delta}$ is a bounded operator and so $\{\bar{\varphi}_t\}$ is bounded in $M^{p}_{2,\delta}$. The induction argument proceeds identically after observing if $\{\bar{\varphi}_{\bar{t}}\}$ is bounded in $M^{p}_{l,\delta}$, then $\{t[M(\bar{\varphi}_t + 1)^{-7} + T(\bar{\varphi}_t + 1)^{-3}]\}$ is bounded in $M^{p}_{k,\delta+2}$. (If s is odd, start the second step at $k = 1$.) Thus $\{\bar{\varphi}_t\}$ is bounded in $M^{p}_{s,\delta}$ and $\{1 + \bar{\varphi}_t\}$ is bounded in $M^{p}_{s,\delta}(1)$.

We now show there is a $C > 0$ such that

$$\left|\frac{d\varphi_t}{dt}\right|_{p,s,\delta} \leqslant C \quad \text{for all } t \in [0,t_0).$$

Using implicit differential differentiation on $\psi(\varphi_t, tM, tT) = 0$ we find $d\varphi_t/dt$ satisfies the equation:

$$\begin{aligned} 8\Delta_g\frac{d\varphi_t}{dt} - (7M\varphi_t^{-8} + 48\pi T\varphi_t^{-4})\frac{d\varphi_t}{dt} \\ = -M\varphi_t^{-7} - 16\pi T\varphi_t^{-3}. \end{aligned}$$

First note $(d\varphi_t/dt)(x) \geqslant 0$ for all $t \in [0,t_0)$ and all $x \in \mathbb{R}^3$. If not, $d\varphi_t/dt$ would have negative minimum. However since

$\varphi_t > 0$ we find, using the above equation, that $\Delta_g(d\varphi_t/dt)$ would be negative in a neighborhood of that minimum. This contradicts Theorem 1.6.

If follows that

$$8\Delta_g \frac{d\varphi_t}{dt} \geqslant -M\varphi_t^{-7} - 16\pi T\varphi_t^{-3}$$
$$\geqslant 8\Delta_g(\Delta_g^{-1}(-M\varphi_t^{-7}/8 - 2\pi T\varphi_t^{-3})),$$

where $\Delta_g^{-1}: M^p_{s-2,\delta+2} \to M^p_{s,\delta}$ is the bounded operator guaranteed by Theorem 1.5 and the closed graph theorem and so

$$8\Delta_g\left(\frac{d\varphi_t}{dt} - \Delta_g^{-1}\left(-\frac{M}{8}\varphi_t^{-7} - 2\pi T\varphi_t^{-3}\right)\right) \leqslant 0.$$

Using the same maximum principle argument as above we find that

$$\frac{d\varphi_t}{dt} - \Delta_g^{-1}\left(-\frac{M}{8}\varphi_t^{-7} - 2\pi T\varphi_t^{-3}\right) \leqslant 0$$

or

$$0 \leqslant \frac{d\varphi_t}{dt} \leqslant \Delta_g^{-1}\left(-\frac{M}{8}\varphi_t^{-7} - 2\pi T\varphi_t^{-3}\right). \tag{6}$$

We know $(-M\varphi_t^{-7}/8 - 2\pi\varphi_t^{-3})$ is uniformly bounded in $M^p_{s-2,\delta+2}$ for $t\in[0,t_0)$. Since Δ_g is a bounded operator we have that $\Delta_g^{-1}(-M\varphi_t^{-7}/8 - 2\pi T\varphi_t^{-3})$ is uniformly bounded in $M^p_{s,\delta}$ for $t\in[0,t_0)$. In particular, it is uniformly bounded in $M^p_{s,\delta}$. It follows from (6) that $d\varphi_t/dt$ is uniformly bounded in $M^p_{0,\delta}$.

We now may use Eq. (5) and the same induction argument found in claim 1, to find $d\varphi_t/dt$ is uniformly bounded in $M^p_{s,\delta}$ for $t\in[0,t_0)$.

Let $t_i \uparrow t_0$. It follows from the Mean Value Theorem that for all i and j, $|\varphi_{t_i} - \varphi_{t_j}|_{P,s,\delta} \leqslant C|t_i - t_j|$. Thus $\{\varphi_{t_i}\}$ is a Cauchy sequence in $M^p_{s,\delta}(1)$ and hence converges to $\varphi_0 \in M^p_{s,\delta}(1)$. By continuity of Ψ we have $\Psi(\varphi_0, t_0 M, t_0 T) = 0$. Also it follows from the maximum principle that $\varphi_0 \geqslant 1$. Hence (4) may be solved at t_0 and this completes the proof of step 2.

Step 3 (Completion of the proof): Let (g,σ,T,ν) be a York data set satisfying condition (I). From step 1 we may find $\alpha \in M^p_{s,\delta}(1)$ such that if we set $g' = \alpha^4 g$, $R(g') = 0$. We now set $\nu' = \alpha^{-10}\nu$, $T' = \alpha^{-8}T$, $\sigma' = \alpha^{-10}\sigma$ and consider the problem

$$8\Delta_{g'}\varphi + M'\varphi^{-7} + 16\pi T'\varphi^{-3} = 0,$$
$$\varphi - 1 \in M^p_{s,\delta},$$
$$\varphi > 0,$$

where $M' = \sigma'\cdot\sigma' = \alpha^{-12}M$. From step 2 this has a unique solution. We now set $\bar{g} = \varphi^4 g'$, $\bar{\pi} = (\varphi\alpha)^{-2}\sigma$, $\bar{\nu} = \varphi^{-10}\nu'$, $\bar{T} = \varphi^{-8}T'$ and obtain the desired initial data.

The smoothness assertion is immediate from the Implicit Function Theorem and Lemma 1.3. Q.E.D.

Remark: It remains an interesting question to determined which $g \in R^p_{s,\delta}$ satisfy condition (I). It is trivially sufficient that $R(g) \geqslant 0$. It follows immediately from the Implicit Function Theorem application in the proof above that it is sufficient that g be sufficiently close to the Euclidean metric in $R^p_{s,\delta}$.

ACKNOWLEDGMENT

The author would like to thank Professor S. Cheng of Princeton University for pointing out the existence of Theorem 1.7.

Note added in proof: A. Chaljub-Simon and Y. Choquet-Bruhot[17] independently proved a theorem similar to Theorem 2.3 of this paper. They work in weighted Holder spaces and use different methods.

[1] Y. Choquet-Bruhat and J.W. York, article in *Einstein Centenary Volume*, edited by Bergman, Goldberg, and Held (Plenum, New York 1979).
[2] J.W. York, article in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge U.P., Cambridge, 1979).
[3] A. Fischer and J. Marsden, article in *Einstein Centenary Volume*, edited by Hawking and Israel (Cambridge U.P., Cambridge, 1979).
[4] J.W. York, J. Math. Phys. **14**, 456–64 (1973).
[5] N. O'Murchadha and J.W. York, Phys. Rev. D **10**, 428–36 (1974).
[6] M. Cantor *et al.*, Commun. Math. Phys. **49**, 197–90 (1976).
[7] J. Marsden and F. Tipler, "Maximal Hypersurfaces and Foliations of Constant Mean Curvature in General Relativity," preprint (1978).
[8] T. Regge and C. Teitelboim, Ann. Phys. (N.Y.) **88**, 286–318 (1974).
[9] M. Cantor, Commun. Math. Phys. **57**, 83–96 (1977).
[10] A. Chaljub-Simon and Y. Choquet-Bruhat, "Equations aux derivees partielles. Solutions asymptoliquement euclidiennes de l'equation de Lichnerowicz," preprint (1978).
[11] M. Cantor, J. Funct. Anal. **18**, 73–84 (1975).
[12] M. Cantor, Compositio Mathematica **38**, Fase. 1, 3–35 (1979).
[13] M. Protter and H. Weinberger, *Maximum Principles in Differential Equations* (Prentice Hall, Englewood Cliffs, New Jersey, 1967).
[14] A.D. Aleksandrov, A.M.S. Trans. Ser. 2 **21**, 341–54 (1962).
[15] N. Dunford and J. Schwartz, *Linear Operators, Part* 1 (Interscience, New York, 1967).
[16] N. O'Murchadha and J.W. York, Phys. Rev. D **10**, 2345–57 (1974).
[17] A. Chaljub-Simon and Y. Choquet-Brujat, Preprint, 1979.

Characteristic surfaces and characteristic initial data for the generalized Einstein–Maxwell field equations

Gregory Walter Horndeski

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
(Received 3 August 1978; revised 14 November 1978)

The characteristic hypersurfaces of the source-free generalized Einstein–Maxwell field equations are investigated. It is shown that such hypersurfaces can be null, and it appears as though they may also be spacelike or timelike. Examples of characteristic initial data on spacelike and timelike hypersurfaces are presented, and it turns out that these examples involve very intense gravitational and electromagnetic fields when the coupling constant k is small in magnitude.

1. INTRODUCTION

In the generalized Einstein–Maxwell field theory the equations governing the symmetric Lorentzian metric tensor g_{ij} and the antisymmetric electromagnetic field tensor F_{ij}, in the absence of sources, are[1]

$$G^{ij} = 8\pi(T^{ij} + kA^{ij}), \tag{1.1}$$

$$F^{ij}{}_{;j} + \tfrac{1}{2}kF_{bc;a}{}^{*}R^{*iabc} = 0, \tag{1.2}$$

and

$$F_{[ab,c]} = 0 \tag{1.3}$$

where

$$T^{ij} := \frac{1}{4\pi}(F^{ia}F^{j}{}_{a} - \tfrac{1}{4}g^{ij}F_{ab}F^{ab}),$$

$$A^{ij} := \frac{1}{8\pi}(F_{al}F_{b}{}^{l}{}^{*}R^{*iajb} + {}^{*}F^{ia}{}_{;b}{}^{*}F^{jb}{}_{;a})$$

and k is a constant with units of (length)^2. These equations are derivable from a variational principle and are uniquely characterized by various physically reasonable assumptions.[2] In addition, when $k = 0$, the above equations reduce to the source-free Einstein–Maxwell field equations.

A *characteristic surface* for the source-free generalized Einstein–Maxwell field equations is a hypersurface S across which g_{ij}, $g_{ij,k}$, and F_{ij} are continuous, but where there are essential discontinuities in either the second derivatives of g_{ij} or the first derivatives of F_{ij} or both. By the discontinuities being "essential" we mean that they cannot be made to disappear by transforming to another chart of the manifold.

It is well known that a characteristic surface for the source-free Einstein–Maxwell field equations must be a null hypersurface.[3] Such hypersurfaces can also serve as characteristic surfaces for the system of Eqs. (1.1)–(1.3). To see this, let us consider the spacetime with metric and electromagnetic field given by[4]

$$ds^2 = L^2(e^{2\beta}dx^2 + e^{-2\beta}dy^2) - 2dudv \tag{1.4}$$

and

$$\tfrac{1}{2}F_{ab}dx^a \wedge dx^b = \Phi_{,x}dx \wedge du + \Phi_{,y}dy \wedge du,$$

where $L = L(u)$, $\beta = \beta(u)$, and $\Phi = \Phi(x,y,u)$. This spacetime will satisfy the source-free generalized Einstein–Maxwell field equations provided

$$-L''L^{-1} - (\beta')^2 = L^{-2}(e^{-2\beta}\Phi_{,x}{}^2 + e^{2\beta}\Phi_{,y}{}^2) + kL^{-4}(\Phi_{,xy}{}^2 - \Phi_{,xx}\Phi_{,yy}) \tag{1.5}$$

and

$$e^{-2\beta}\Phi_{,xx} + e^{2\beta}\Phi_{,yy} = 0, \tag{1.6}$$

with $' := \partial/\partial u$. If $k > 0$, then we cannot produce a nontrivial solution to Eqs. (1.5) and (1.6) which has

$$L''L^{-1} + (\beta')^2 = 0. \tag{1.7}$$

However, if $k < 0$, then we can construct such solutions[5] by taking any solution to (1.7) and defining

$$\Phi_{,x} = Le^{\beta}\mathrm{Im}(C) \quad \text{and} \quad \Phi_{,y} = Le^{-\beta}\mathrm{Re}(C),$$

where the complex valued function C is given by

$$C = B\exp[Ll^{-1}\exp(i\psi)(e^{\beta}x + ie^{-\beta}y)]$$

with B and ψ being arbitrary complex and real valued functions of u resp., and $-l^2 = k$. Since the hypersurfaces $u = \text{const}$ are null hypersurfaces of the spacetime with metric (1.4), we see that if we choose β to be of class C^2 everywhere, except along $u = 0$, where it is of class C^1, then when $k < 0$ we can find solutions to the source-free generalized Einstein–Maxwell field equations which have null hypersurfaces acting as characteristic surfaces.

The purpose of this paper is to argue that timelike and spacelike hypersurfaces can also act as characteristic surfaces for the source-free generalized Einstein–Maxwell field equations. Due to the intimate relationship between characteristic hypersurfaces and "wave fronts," this observation leads one to suspect that in the generalized Einstein–Maxwell field theory it may be possible to propagate information at speeds greater than light, more exactly, along paths with spacelike tangent vector. If this were the case, then it could be argued that the generalized Einstein–Maxwell field theory should be dismissed as being unphysical. However, it is best to put off judgement on this issue until after examining those conditions which may lead to "acausal propagation." The general form of these conditions will be derived in the next section, and in Sec. 3 particular cases will be investigated.

 0022-2488/79/081745-07$01.00

2. CONSTRAINT EQUATIONS AND THE CHARACTERISTIC MATRIX

Let S be a hypersurface in the four-dimensional manifold M. We require S to be regular in the sense that the manifold topology on S is the same as its topology as a subset of M. Thus S cannot come "arbitrarily close to itself." If $p \in S$, then there exists a chart $x = (x^0, x^1, x^2, x^3)$ at p with domain U which is such that in U, S is given by $x^0 = 0$. We assume that in U

$$g_{ij};\ g_{ij,k};\ g_{ij,k\alpha};\ F_{ij};\ F_{ij,\alpha} \tag{2.1}$$

are all continuous across S and that the only second order derivatives of g_{ij} and first order derivatives of F_{ij} which can experience discontinuities across S are

$$g_{ij,00} \quad \text{and} \quad F_{ij,0}. \tag{2.2}$$

We shall now determine those conditions which the functions presented in (2.1) must satisfy on $S_u := S \cap U$ in order to guarantee that some of the functions given in Eq. (2.2) cannot be specified on S_u through use of the field equations (1.1)–(1.3). It is only when these conditions are satisfied that $g_{ij,00}$ and $F_{ij,0}$ can be discontinuous across the hypersurface S.

In order to simplify the form of the following expressions, we shall let the symbol $[C]$ denote continuous quantities built from the functions presented in Eq. (2.1).

We begin our construction of the aforementioned conditions with Eq. (1.3). This equation tells us that

$$F_{\alpha\beta,0} = [C]_{\alpha\beta}, \tag{2.3}$$

and hence $F_{\alpha\beta,0}$ is continuous across S. Thus the only x^0 derivative of F_{ij} which may be discontinuous across S is $F_{0\alpha,0}$.

Since

$$*R^{*abcd} := \frac{-1}{4g}\epsilon^{abrs}\epsilon^{cdtu}R_{rstu},$$

where $g := \det(g_{ab})$, we find that

$$*R^{*0\alpha0\beta} = [C]^{\alpha\beta}, \quad *R^{*0\alpha\beta\gamma} = [C]^{\alpha\beta\gamma} \tag{2.4}$$

and

$$*R^{*\alpha\beta\gamma\delta} = \frac{1}{2g}\epsilon^{0\alpha\beta\rho}\epsilon^{0\gamma\delta\tau}g_{\rho\tau,00} + [C]^{\alpha\beta\gamma\delta}.$$

Using Eqs. (2.3) and (2.4), we discover that

$$G^{0j} - 8\pi(T^{0j} + kA^{0j})$$

and

$$F^{0j}{}_{;j} + \tfrac{1}{2}kF_{bc;a}{}^{*}R^{*0abc}$$

are built solely from the functions given in Eq. (2.1). Consequently the six field equations

$$\begin{aligned} &G^{0j} = 8\pi(T^{0j} + kA^{0j}),\\ &F^{0j}{}_{;j} + \tfrac{1}{2}kF_{bc;a}{}^{*}R^{*0abc} = 0,\\ &F_{[\alpha\beta,\gamma]} = 0 \end{aligned} \tag{2.5}$$

simply represent constraints upon the functions (2.1). We shall now employ the nine remaining field equations in (1.1) and (1.2) to derive a linear system of nine equations for $g_{\alpha\beta,00}$ and $F_{0\alpha,0}$ in terms of the functions (2.1). The 9×9 matrix appearing in this system of linear equations is called the characteristic matrix for the source-free generalized Einstein–Maxwell field equations. It will turn out that a solution to Eqs. (1.1)–(1.3) will have S as a characteristic surface if and only if it yields a characteristic matrix which has a vanishing determinant on S.

A straightforward calculation involving Eqs. (2.3) and (2.4) shows that

$$\begin{aligned} &F^{\gamma j}{}_{;j} = (g^{\gamma 0}g^{\mu 0} - g^{\mu\gamma}g^{00})F_{0\mu,0} + [C]^{\gamma},\\ &F_{bc;a}{}^{*}R^{*\gamma abc} = \tfrac{1}{2}M^{\mu\nu,\gamma}g_{\mu\nu,00} + 2\,{}^{*}R^{*0\gamma\mu0}F_{0\mu,0} + [C]^{\gamma},\\ &G^{\alpha\beta} = G^{\alpha\beta\mu\nu}g_{\mu\nu,00} + [C]^{\alpha\beta}, \end{aligned}$$

and

$$8\pi(T^{\alpha\beta} + kA^{\alpha\beta}) = \frac{k}{2g}F_{\rho l}F_{\zeta}{}^{l}\epsilon^{0\alpha\rho\mu}\epsilon^{0\beta\zeta\nu}g_{\mu\nu,00} - kM^{\alpha\beta,\mu}F_{0\mu,0} + [C]^{\alpha\beta},$$

where

$$M^{\alpha\beta,\mu} := \frac{1}{2g}(\epsilon^{0\xi\delta\alpha}\epsilon^{0\mu\zeta\beta} + \epsilon^{0\xi\delta\beta}\epsilon^{0\mu\zeta\alpha})F_{\xi\delta;\zeta} \tag{2.6}$$

and

$$\begin{aligned} G^{\alpha\beta\mu\nu} := \tfrac{1}{2}\{&g^{\alpha0}g^{\beta(\mu}g^{\nu)0} + g^{\beta0}g^{\alpha(\mu}g^{\nu)0} - g^{\alpha0}g^{\beta0}g^{\mu\nu}\\ &- g^{\alpha(\mu}g^{\nu)\beta}g^{00} - g^{\alpha\beta}g^{0\mu}g^{0\nu} + g^{\alpha\beta}g^{00}g^{\mu\nu}\}. \end{aligned} \tag{2.7}$$

Consequently, when $(i,j) = (\alpha,\beta)$ and $i = \gamma$ in Eqs. (1.1) and (1.2) respectively, we obtain the following system of linear equations for $g_{\mu\nu,00}$ and $F_{0\mu,0}$:

$$N^{\alpha\beta,\mu\nu}g_{\mu\nu,00} + kM^{\alpha\beta,\mu}F_{0\mu,0} = [C]^{\alpha\beta}, \tag{2.8}$$

$$\tfrac{1}{4}k\,M^{\mu\nu,\gamma}g_{\mu\nu,00} + P^{\gamma\mu}F_{0\mu,0} = [C]^{\gamma}, \tag{2.9}$$

where

$$N^{\alpha\beta,\mu\nu} := G^{\alpha\beta\mu\nu} - \frac{k}{2g}F_{\rho l}F_{\zeta}{}^{l}\epsilon^{0\alpha\rho(\mu}\epsilon^{\nu)\zeta\beta0} \tag{2.10}$$

and

$$P^{\gamma\mu} := g^{\gamma0}g^{\mu0} - g^{\gamma\mu}g^{00} + k\,{}^{*}R^{*0\gamma\mu0}. \tag{2.11}$$

In order to reformulate Eqs. (2.8) and (2.9) in matrix language, we shall let capital Latin indices range from 1 to 6 and identify these indices with the index pairs $\mu\nu$ in accordance with the following scheme:

$$\begin{aligned} &1 := (11), \quad 2 := (12), \quad 3 := (13),\\ &4 := (22), \quad 5 := (23), \quad 6 := (33). \end{aligned}$$

In addition a tilde will be placed above any quantity using capital Latin indices which has been built from a quantity involving lower case Greek indices. Thus, e.g.,

$$\widetilde{N}^{3,5} := N^{13,23}, \quad \text{and} \quad \widetilde{M}^{4,2} := M^{22,2}.$$

This notation permits us to rewrite Eqs. (2.8) and (2.9) as follows:

$$\begin{pmatrix} \chi^{AU} & k\widetilde{M}^{A,\mu} \\ \tfrac{1}{4}k\omega^{\gamma U} & P^{\gamma\mu} \end{pmatrix}\begin{pmatrix} g_{U,00} \\ F_{0\mu,0} \end{pmatrix} = \begin{pmatrix} [C]^{A} \\ [C]^{\gamma} \end{pmatrix}, \tag{2.12}$$

where

$$[\chi^{AU}]:=\begin{bmatrix} \widetilde{N}^{1,1} & 2\widetilde{N}^{1,2} & 2\widetilde{N}^{1,3} & \widetilde{N}^{1,4} & 2\widetilde{N}^{1,5} & \widetilde{N}^{1,6} \\ \widetilde{N}^{2,1} & 2\widetilde{N}^{2,2} & 2\widetilde{N}^{2,3} & \widetilde{N}^{2,4} & 2\widetilde{N}^{2,5} & \widetilde{N}^{2,6} \\ \vdots & - & - & - & - & \vdots \\ \widetilde{N}^{6,1} & 2\widetilde{N}^{6,2} & 2\widetilde{N}^{6,3} & \widetilde{N}^{6,4} & 2\widetilde{N}^{6,5} & \widetilde{N}^{6,6} \end{bmatrix} \tag{2.13}$$

and

$$[\omega^{\gamma U}]:=\begin{bmatrix} \widetilde{M}^{1,1} & 2\widetilde{M}^{2,1} & 2\widetilde{M}^{3,1} & \widetilde{M}^{4,1} & 2\widetilde{M}^{5,1} & \widetilde{M}^{6,1} \\ \widetilde{M}^{1,2} & 2\widetilde{M}^{2,2} & 2\widetilde{M}^{3,2} & \widetilde{M}^{4,2} & 2\widetilde{M}^{5,2} & \widetilde{M}^{6,2} \\ \widetilde{M}^{1,3} & 2\widetilde{M}^{2,3} & 2\widetilde{M}^{3,3} & \widetilde{M}^{4,3} & 2\widetilde{M}^{5,3} & \widetilde{M}^{6,3} \end{bmatrix} \tag{2.14}$$

The 9×9 matrix appearing on the left-hand side of Eq. (2.12) will be denoted by Q and is called the *characteristic matrix* for the source-free generalized Einstein–Maxwell field equations. Due to Eqs. (2.4), (2.6), (2.10), (2.11), (2.13), and (2.14) it is clear that the coefficients of this matrix are built only from the functions given in Eq. (2.1).

Since Eqs. (2.3), (2.5), and (2.12) embody all of the restrictions imposed by the source-free generalized Einstein–Maxwell field equations we see (as usual) that these equations do not suffice to determine all of the second derivatives given in Eq. (2.2) on S (since $g_{0a,00}$ does not appear in any of these equations). This ambiguity can be removed by choosing our original chart x at $p\in S$ to be a Gaussian coordinate system (see pp. 35–37 of Ref. 3). In terms of such a chart we have S given by $x^0=0$ in U with $g_{00}=\epsilon(=\pm 1)$ and $g_{0\alpha,0}=0$ on U. As a result of using such coordinates, $g_{0a,00}=0$, and thus we can employ Eq. (2.3) to deduce that the only second derivatives which need to be determined on S_u are $g_{\alpha\beta,00}$ and $F_{0\alpha,0}$. These derivatives will be uniquely determined on S_u by Eq. (2.12) provided that the characteristic matrix Q is nonsingular on S_u. Hence we can now conclude that a solution to the source-free generalized Einstein–Maxwell field equations (in terms of the Gaussian coordinates) will have S as a characteristic surface if and only if the associated characteristic matrix is singular on S.

The problem of interpreting the condition $\det Q=0$ in general appears to be quite intractible. Thus in the next section we shall investigate various restrictions on the functions presented in Eq. (2.1) which lead to a singular characteristic matrix.

We shall now consider the effect that a coordinate transformation of the form

$$\bar{x}^0=x^0,\quad \bar{x}^\alpha=\bar{x}^\alpha(x^\beta) \tag{2.15}$$

has on the equation $\det Q=0$. To begin with, it is clear that under (2.15), $M^{\alpha\beta,\gamma}$, $N^{\alpha\beta,\mu\nu}$, and $P^{\gamma\mu}$ transform as 3-tensors; e.g.,

$$\bar{P}^{\gamma\mu}=P^{\sigma\tau}\bar{B}^{\gamma}_{\ \sigma}\bar{B}^{\mu}_{\ \tau},$$

where $\bar{B}^{\gamma}_{\ \sigma}:=\partial\bar{x}^\gamma/\partial x^\sigma$. This in turn implies that there exists an invertible 6×6 matrix $(\bar{X}^{A}_{\ U})$, which is such that

$$\bar{\widetilde{N}}^{A,U}=\widetilde{N}^{B,V}\bar{X}^{A}_{\ B}\bar{X}^{U}_{\ V}$$

and

$$\bar{\widetilde{M}}^{A,\mu}=\widetilde{M}^{B,\tau}\bar{X}^{A}_{\ B}\bar{B}^{\mu}_{\ \tau}.$$

As a result of these facts

$$\det(\bar{\chi}^{AU})=\det(\chi^{BV})(\det\bar{X}^{A}_{\ B})^2,$$

$$\det(\bar{P}^{\gamma\mu})=\det(P^{\sigma\tau})(\det\bar{B}^{\gamma}_{\ \sigma})^2,$$

and

$$\det\bar{Q}=\det Q\,(\det\bar{X}^{A}_{\ B})^2(\det\bar{B}^{\gamma}_{\ \sigma})^2, \tag{2.16}$$

where it can be shown that $\det\bar{X}^{A}_{\ B}=(\det\bar{B}^{\gamma}_{\ \sigma})^4$. Consequently, $\det Q=0$ if and only if $\det\bar{Q}=0$; with a similar remark applying to $\det(\chi^{AU})=0$ and $\det(P^{\gamma\mu})=0$. These observations will prove to be very useful in the next section. In passing one should note that the constraint equations are satisfied in the barred coordinate system if and only if they are satisfied in the unbarred coordinate system.

The remainder of this section will be devoted to a brief discussion of the Cauchy problem for the source-free generalized Einstein–Maxwell field equations.

Let S be a hypersurface in a four-dimensional manifold M and let x be a chart of M at $p\in S$ with domain U which is such that in U, S is given by $x^0=0$. If we require that x be a Gaussian normal coordinate system, then the Cauchy problem for the source-free generalized Einstein–Maxwell field equations involves specifying the values of $g_{\alpha\beta}$, $g_{\alpha\beta,0}$, and F_{ij} on $S_u:=S\cap U$, and then determining a solution to Eqs. (1.1)–(1.3) for g_{ij} and F_{ij} on a neighborhood W of S_u, which agrees with the prescribed initial data, and is such that $g_{0\alpha}=0$ and $g_{00}=\pm 1$ on W (-1 if S is to be spacelike, $+1$ if S is to be timelike). Due to our work in this section we know that the initial data on S_u cannot be arbitrarily specified since it must satisfy the constraint equation (2.5) on S_u. [Note that a knowledge of g_{ij}, $g_{ij,0}$, and F_{ij} on S serves to determine all of the functions given in Eq. (2.1) on S.] If the initial data has been chosen to satisfy this equation, and consists of analytic functions of the x^α's which are such that $\det Q\neq 0$ on S_u, then we shall show that the Cauchy problem (in terms of Gaussian normal coordinates) can be solved with g_{ij} and F_{ij} being analytic functions of the x^i's. We begin our proof of this fact with a few remarks concerning Eqs. (1.1)–(1.3).

Since the tensor field $F^{ij}_{\ \ ;j}+\frac{1}{2}kF_{bc;a}{}^*R^{*iabc}$ is identically divergence-free, it is easy to see that if we have a solution to

$$F^{\alpha j}_{\ \ ;j}+\tfrac{1}{2}kF_{bc;a}{}^*R^{*\alpha abc}=0$$

and

$$F_{[0\alpha,\beta]}=0$$

for g_{ij} and F_{ij}, on a connected neighborhood $W\subset U$ of S_u, which is such that

$$F^{0j}_{\ \ ;j}+\tfrac{1}{2}kF_{bc;a}{}^*R^{*0abc}=0$$

and

$$F_{[\alpha\beta,\gamma]}=0$$

on S_u, then Eqs. (1.2) and (1.3) are satisfied on[6] W.

It can be shown[7] that the tensor field $G^{ij}-8\pi(T^{ij}+kA^{ij})$ is divergence-free when Eqs. (1.2) and (1.3) are satisfied. Using this result, along with the above observation concerning Eqs. (1.2) and (1.3), we can prove

the following proposition on the Cauchy problem for the source-free generalized Einstein–Maxwell field equations.

Proposition 1: *If the initial data on S_u has been chosen to satisfy the constraint equation* (2.5), *and we have produced an analytic solution to the equations*

$$G^{\alpha\beta} = 8\pi(T^{\alpha\beta} + kA^{\alpha\beta}),$$
$$F^{\alpha j}{}_{;j} + \tfrac{1}{2}kF_{bc;a}{}^{*}R^{*\alpha abc} = 0, \tag{2.17}$$
$$F_{[0\alpha,\beta]} = 0,$$

in terms of Gaussian normal coordinates on a connected neighborhood $W \subset U$ of S_u, which argees with the initial data, then this solution is also a solution to the system of partial differential equations (1.1)–(1.3) *on W.*

The similarity between Proposition 1 and the corresponding result for the Einstein–Maxwell field theory is remarkable in view of the complexity of the generalized Einstein–Maxwell field equations.

Equation (2.17) represents a system of twelve equations for the twelve unknown functions $g_{\alpha\beta}$ and F_{ij}. Due to our previous work we know that on S_u Eq. (2.17) can be solved for $g_{\alpha\beta,00}$ and $F_{ij,0}$ provided that the characteristic matrix Q is nonsingular. Thus we can now appeal to the Cauchy–Kowalewsky theorem[8] to conclude that if we are given analytic initial data on S_u, which is such that $\det Q \neq 0$ on S_u, then there exists an analytic solution to Eq. (2.17) for $g_{\alpha\beta}$ and F_{ij} on a connected neighborhood W of S_u. Upon combining this result with Proposition 1 we obtain the following:

Proposition 2: *If the initial data on S_u has been chosen to satisfy the constraint equation* (2.5), *and consists of analytic functions of the x^{α}'s which are such that $\det Q \neq 0$ on S_u, then on a connected neighborhood of S_u there exists an analytic solution to the source-free generalized Einstein–Maxwell field equations, which agrees with the initial data, and has x as a Gaussian normal coordinate system.*

3. CHARACTERISTIC INITIAL DATA

In the Introduction we saw that null hypersurfaces can serve as characteristic surfaces for the source-free generalized Einstein–Maxwell field equations. The purpose of this section is to present a fairly strong argument for the case that timelike and spacelike hypersurfaces can also act as characteristic surfaces for these field equations. To that end, let S be an open subset of the hypersurface $x^0 = 0$ in $\mathbb{R}^4$ where $x = (x^0, x^1, x^2, x^3)$ is the standard chart of $\mathbb{R}^4$. As initial data on S for the source-free generalized Einstein–Maxwell field equations we choose to specify g_{ij}, $g_{ij,0}$, and F_{ij} as differentiable functions on S with $g_{00} = \epsilon(= \pm 1)$, $g_{0\alpha} = 0$, and $g_{0\alpha,0} = 0$. This data will be called *characteristic initial data* (in a Gaussian normal coordinate system) if it satisfies the constraint equation (2.5) and yields a singular characteristic matrix Q on S. (Note that when evaluating the constraint equations the field equations $F_{[0\alpha,\beta]} = 0$ are used to eliminate terms involving $F_{\alpha\beta,0}$.) We shall show that it is possible to present characteristic initial data on S and hence the source-free generalized Einstein–Maxwell field equations may admit spacelike and timelike characteristic surfaces.

In order to determine the required characteristic initial data, we note that if $M^{\alpha\beta,\gamma} = 0$ on S then the characteristic matrix Q assumes the block form [cf. Eq. (2.12) and (2.14)]

$$Q = \begin{pmatrix} \chi^{AU} & 0 \\ 0 & P^{\gamma\mu} \end{pmatrix}. \tag{3.1}$$

In this case $\det Q = 0$ if and only if either $\det(\chi^{AU}) = 0$ or $\det(P^{\gamma\mu}) = 0$. We shall now examine what conditions the initial data must satisfy if $\det(\chi^{AU}) = 0$.

Suppose that S is spacelike; i.e., $g_{00} = -1$ on S. If $p \epsilon S$, then we can perform a coordinate transformation on the form (2.15) to arrange that $(\bar{g}_{ij}) = \mathrm{diag}(-1,1,1,1)$ at p. A lengthy calculation shows that at p

$$\begin{aligned}
\det(\bar{\chi}^{AU}) = \tfrac{-1}{32}\{&1 - 2(A + D + F) + (A^2 + D^2 + F^2 \\
&- 2B^2 - 2C^2 - 2E^2 + 4AD + 4AF + 4DF) - 2(A^2D \\
&+ A^2F + AD^2 + D^2F + AF^2 + DF^2 - AB^2 - B^2D - AC^2 \\
&- C^2F - DE^2 - E^2F - 2B^2F - 2C^2D - 2AE^2 + 2BCE \\
&+ 4ADF) + (B^4 + C^4 + E^4 + A^2D^2 + A^2F^2 + D^2F^2 + 2B^2C^2 \\
&+ 2B^2E^2 + 2C^2E^2 - 2B^2F^2 - 2C^2D^2 - 2A^2E^2 - 2AB^2D \\
&- 4AB^2F - 4B^2DF - 4AC^2D - 2AC^2F - 4C^2DF \\
&- 4ADE^2 - 4AE^2F - 2DE^2F + 4A^2DF + 4AD^2F \\
&+ 4ADF^2 + 4ABCE + 4BCDE + 4BCEF) - 2(B^4F \\
&+ C^4D + AE^4 - AB^2F^2 - B^2DF^2 - AC^2D^2 - C^2D^2F \\
&- A^2DE^2 - A^2E^2F + B^2C^2D + B^2C^2F + AB^2E^2 \\
&+ B^2E^2F + C^2DE^2 + AC^2E^2 + A^2D^2F + A^2DF^2 \\
&+ AD^2F^2 - 2AB^2DF - 2AC^2DF - 2ADE^2F - 2B^3CE \\
&- 2BC^3E - 2BCE^3 + 2ABCDE + 2ABCEF + 2BCDEF) \\
&+ (B^4F^2 + C^4D^2 + A^2E^4 + 2B^2C^2DF + 2AB^2E^2F \\
&+ 2AC^2DE^2 - 2AB^2DF^2 - 2AC^2D^2F - 2A^2DE^2F \\
&- 4B^3CEF - 4BC^3DE - 4ABCE^3 + 4B^2C^2E^2 \\
&+ A^2D^2F^2 + 4ABCDEF)\},
\end{aligned} \tag{3.2}$$

where

$$A := k\bar{F}_{1h}\bar{F}_1{}^h, \quad B := k\bar{F}_{1h}\bar{F}_2{}^h, \quad C := k\bar{F}_{1h}\bar{F}_3{}^h,$$
$$D := k\bar{F}_{2h}\bar{F}_2{}^h, \quad E := k\bar{F}_{2h}\bar{F}_3{}^h, \quad F := k\bar{F}_{3h}\bar{F}_3{}^h.$$

Recall that if N is a field of unit normals for a nonnull hypersurface Σ, then on Σ we can decompose F_{ij} as follows:

$$F_{ij} = -\epsilon[2N_{[i}E_{j]} + (-g)^{1/2}\epsilon_{ijab}N^a B^b], \tag{3.3}$$

where

$$\epsilon := N_a N^a, \quad E_a := F_{ab}N^b, \quad B_a := -{}^{*}F_{ab}N^b. \tag{3.4}$$

The vectors E^a and B^a are called the *electric* and *magnetic field vectors* (on Σ). F_{ij} is said to be *pure electric* (*pure magnetic* resp.) on Σ if $B^a = 0$ ($E^a = 0$, resp.). When $\Sigma = S$, we choose $N = \partial/\partial x^0$ restricted to S.

Using Eqs. (2.16) and (3.2), it can be shown that if S is spacelike then at p

$$\det(\chi^{AU})$$

$$= \begin{cases} \dfrac{-1}{32g^4}(1 - kB_aB^a)^4 & \text{if } F_{ij} \text{ is pure magnetic on } S, \\ \dfrac{-1}{32g^4}(1 + kE_aE^a)^2 & \text{if } F_{ij} \text{ is pure electric on } S. \end{cases} \tag{3.5}$$

Since Eq. (3.4) implies that E^a and B^a are tangent to S we see that when S is spacelike B^aB_a and E^aE_a must be nonnegative. Thus we can now use Eq. (3.5), along with the fact that our choice of the point $p \in S$ was arbitrary, to deduce the following:

Lemma 1: *If the initial data is such that*

(i) $M^{\alpha\beta,\gamma} = 0$ *on* S;

(ii) S *is spacelike; and*

(iii) Either $k > 0$, *and* F_{ij} *is pure magnetic with* $1 = kB_aB^a$, *or* $k < 0$, *and* F_{ij} *is pure electric with* $-1 = kE_aE^a$, *then* $\det Q = 0$ *on* S.

When S is timelike we can choose our chart $\bar{x}$ at $p \in S$ so that $(\bar{g}_{ij}) = \mathrm{diag}(1,1,1,-1)$ at p. In terms of this chart $\det(\chi^{AU})$ has the same terms as those appearing in Eq. (3.2), but many of the signs are changed. The end result of the calculation is that when S is timelike then at p

$$\det(\chi^{AU}) = \begin{cases} \dfrac{-1}{32g^4}(1 + kB_aB^a)^4 & \text{if } F_{ij} \text{ is pure magnetic on } S, \\ \dfrac{-1}{32g^4}(1 - kE_aE^a)^2 & \text{if } F_{ij} \text{ is pure electric on } S. \end{cases}$$

In this case B^a and E^a can be either spacelike, timelike, or null, and hence the vanishing $\det(\chi^{AU})$ when F_{ij} is pure magnetic or pure electric places no restriction upon the sign of k. Consequently, we have the following:

Lemma 2: *If the initial data is such that*

(i) $M^{\alpha\beta,\gamma} = 0$ *on* S;

(ii) S *is timelike; and*

(iii) F_{ij} *is either pure magnetic with* $-1 = kB_aB^a$ *or pure electric with* $1 = kE_aE^a$, *then* $\det Q = 0$ *on* S.

Lemmas 1 and 2 both require that $M^{\alpha\beta,\gamma} = 0$ on S. It is an elementary matter to prove the following two lemmas concerning this condition.

Lemma 3: *If* S *is nonnull and* F_{ij} *is pure magnetic on* S, *then* $M^{\alpha\beta,\gamma} = 0$ *and* $F_{[\alpha\beta,\gamma]} = 0$ *on* S *if and only if* $F_{\alpha\beta|\gamma} = 0$, *where a vertical bar denotes covariant differentiation with respect to the three-dimensional metric* $\bar{g}_{\alpha\beta}$ *on* S *given by* $\bar{g}_{\alpha\beta} := g_{\alpha\beta}$.

Lemma 4: *If* S *is nonnull and* F_{ij} *is pure electric on* S, *with* $F_{ij} \neq 0$, *then* $M^{\alpha\beta,\gamma} = 0$ *on* S *if and only if* $\Omega_{\alpha\beta} = \Lambda E_\alpha E_\beta$, *where* Λ *is a differentiable function on* S *and* $\Omega_{\alpha\beta} := -\frac{1}{2}g_{\alpha\beta,0}$.

If F_{ij} is pure magnetic on S, and $F_{\alpha\beta|\gamma} = 0$, then the vector field B^α must be a parallel vector field of the pseudo-Riemannian space $V_3 = (S, \bar{g}_{\alpha\beta})$, and hence $B^\alpha \bar{R}_{\alpha\beta\mu\nu} = 0$. Using this fact along with the fact that the Weyl tensor of a V_3 vanishes identically, we can show that

$$\bar{R}_{\alpha\beta\mu\nu} = \frac{\bar{R}}{2}(\bar{g}_{\alpha\mu}\bar{g}_{\beta\nu} - \bar{g}_{\alpha\nu}\bar{g}_{\beta\mu}) + \frac{\bar{R}}{2B^\lambda B_\lambda}(B_\alpha B_\nu \bar{g}_{\beta\mu} - B_\alpha B_\mu \bar{g}_{\beta\nu} + B_\beta B_\mu \bar{g}_{\alpha\nu} - B_\beta B_\nu \bar{g}_{\alpha\mu}), \tag{3.6}$$

provided $B^\lambda B_\lambda \neq 0$. In addition, since $\bar{G}^{\alpha\beta}{}_{|\beta} = 0$, we must have

$$B^\alpha \bar{R}_{|\alpha} = 0.$$

Our work so far has shown us how to choose initial data so that $\det Q = 0$ on S. However, not all such initial data will satisfy the constraint equation (2.5), which we shall now proceed to investigate.

If S is spacelike, and F_{ij} is pure magnetic on S with $F_{\alpha\beta|\gamma} = 0$, then the constraint equation (2.5) becomes

$$\begin{aligned} &kF_{\mu\beta}\Omega^\mu{}_\gamma {}^*R^{*0\beta 0\gamma} = 0, \\ &\tfrac{1}{2}(\bar{R} + \Omega^\alpha{}_\alpha\Omega^\beta{}_\beta - \Omega^\alpha{}_\beta\Omega^\beta{}_\alpha) \\ &\quad = B^\mu B_\mu + kF_{\beta\lambda}F^\lambda{}_\gamma {}^*R^{*0\beta 0\gamma}, \\ &\Omega_\alpha{}^\lambda{}_{|\lambda} - \Omega_\lambda{}^\lambda{}_{|\alpha} = kF_{\mu\lambda}F^{\nu\lambda}\delta^{\mu\rho\xi}_{\alpha\nu\tau}\Omega_\xi{}^\tau{}_{|\rho}, \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} {}^*R^{*0\beta 0\gamma} &= g_{00}\bar{G}^{\beta\gamma} + \Omega^{\beta\sigma}\Omega^\gamma{}_\sigma - \Omega^{\beta\gamma}\Omega^\sigma{}_\sigma \\ &\quad + \tfrac{1}{2}\bar{g}^{\beta\gamma}(\Omega^\tau{}_\tau\Omega^\sigma{}_\sigma - \Omega^\tau{}_\sigma\Omega^\sigma{}_\tau) \end{aligned} \tag{3.8}$$

and lower case Greek indices have been raised and lowered with $\bar{g}^{\alpha\beta}$ and $\bar{g}_{\alpha\beta}$ resp. When S is spacelike, and F_{ij} is pure electric with $\Omega_{\alpha\beta} = \Lambda E_\alpha E_\beta$, then Eq. (2.5) becomes

$$\begin{aligned} &E^\alpha{}_{|\alpha} + k\bar{G}^{\alpha\beta}E_{\alpha|\beta} = 0, \\ &\tfrac{1}{2}\bar{R} = E^\alpha E_\alpha + k\bar{G}^{\alpha\beta}E_\alpha E_\beta, \\ &\Omega_\alpha{}^\lambda{}_{|\lambda} - \Omega_\lambda{}^\lambda{}_{|\alpha} = 0. \end{aligned} \tag{3.9}$$

Thus the problem of producing characteristic initial data on a spacelike hypersurface for the source-free generalized Einstein–Maxwell field equations has reduced to finding solutions to either Eq. (3.7) or (3.9) which lead to $\det Q = 0$ in the manner described in Lemma 1. Upon noting Eq. (3.6) and the fact that when F_{ij} is pure magnetic $F_{\alpha\beta|\gamma} = 0$ if and only if B^α is parallel [which follows from Eq. (3.3)], we get the following sets of characteristic initial data.[9]

Set I: *Characteristic initial data when* S *is spacelike and* $k > 0$.

(i) F_{ij} is pure magnetic with $1 = kB_aB^a$;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S, \bar{g}_{\alpha\beta})$ admits B^α as a parallel vector field and has constant scalar curvature $\bar{R}$ with $k^{-1} - \frac{1}{2}\bar{R} \geqslant 0$; and

(iii) $\Omega_{\alpha\beta} = \Lambda\bar{g}_{\alpha\beta}$, where $\Lambda^2 = k^{-1} - \frac{1}{2}\bar{R}$.

Set II: *Characteristic initial data when* S *is spacelike and* $k > 0$.

(i) F_{ij} is pure magnetic with $1 = kB_aB^a$;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S, \bar{g}_{\alpha\beta})$ admits B^α as a parallel vector field and has constant scalar curvature $\bar{R} = 2k^{-1}$; and

(iii) $\Omega_{\alpha\beta} = \Lambda B_\alpha B_\beta$, where Λ is an arbitrary differentiable function on S.

Set III: *Characteristic initial data when* S *is spacelike and* $k < 0$.

(i) F_{ij} is pure electric with $-1 = kE^{a}E_{a}$ and $E^{\alpha}{}_{|\alpha} = 0$;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S,\bar{g}_{\alpha\beta})$ is a space of constant curvature[10] with scalar curvature $\bar{R} = -3k^{-1}$, and

(iii) $\Omega_{\alpha\beta} = 0$.

We shall now turn our attention to the case in which S is timelike.

When S is timelike, and the electromagnetic field is pure magnetic with $F_{\alpha\beta|\gamma} = 0$, then the constraint equation (2.5) becomes

$$kF_{\mu\beta}\Omega^{\mu}{}_{\gamma}{}^{*}R^{*0\beta 0\gamma} = 0,$$

$$-\tfrac{1}{2}(\bar{R} - \Omega^{\alpha}{}_{\alpha}\Omega^{\beta}{}_{\beta} + \Omega^{\alpha}{}_{\beta}\Omega^{\beta}{}_{\alpha}) = B^{\mu}B_{\mu} + kF_{\beta\lambda}F_{\gamma}{}^{\lambda}{}^{*}R^{*0\beta 0\gamma}, \tag{3.10}$$

$$\Omega_{\alpha}{}^{\lambda}{}_{|\lambda} - \Omega^{\lambda}{}_{\lambda|\alpha} = kF_{\mu\lambda}F^{\nu\lambda}\delta^{\mu\rho\zeta}_{\alpha\nu\tau}\Omega_{\zeta}{}^{\tau}{}_{|\rho},$$

where ${}^{*}R^{*0\beta 0\gamma}$ is defined by Eq. (3.8).

If S is timelike, and F_{ij} is pure electric with $\Omega_{\alpha\beta} = \Lambda E_{\alpha}E_{\beta}$, then Eq. (2.5) becomes

$$E^{\alpha}{}_{|\alpha} + k\bar{G}^{\alpha\beta}E_{\alpha|\beta} = 0,$$

$$-\tfrac{1}{2}\bar{R} = E^{\alpha}E_{\alpha} + k\bar{G}^{\alpha\beta}E_{\alpha}E_{\beta},$$

$$\Omega_{\alpha}{}^{\lambda}{}_{|\lambda} - \Omega_{\lambda}{}^{\lambda}{}_{|\alpha} = 0. \tag{3.11}$$

Upon combining Eqs. (3.10) and (3.11) with Lemmas 2, 3, and 4, we obtain the following three sets of characteristic initial data.

Set IV: Characteristic initial data when S is timelike.

(i) F_{ij} is pure magnetic with $-1 = kB_{a}B^{a}$;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S,\bar{g}_{\alpha\beta})$ admits B^{α} as a parallel vector field and has constant scalar curvature $\bar{R}$ with $\frac{1}{2}\bar{R} - k^{-1} \geqslant 0$; and

(iii) $\Omega_{\alpha\beta} = \Lambda\bar{g}_{\alpha\beta}$, where $\Lambda^{2} = \frac{1}{2}\bar{R} - k^{-1}$.

Set V. Characteristic initial data when S is timelike.

(i) F_{ij} is pure magnetic with $-1 = kB_{a}B^{a}$;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S,\bar{g}_{\alpha\beta})$ admits B^{α} as a parallel vector field and has constant scalar curvature $\bar{R} = 2k^{-1}$; and

(iii) $\Omega_{\alpha\beta} = \Lambda B_{\alpha}B_{\beta}$, where Λ is any differentiable function on S.

Set VI. Characteristic initial data when S is timelike.

(i) F_{ij} is pure electric on S with $1 = kE_{a}E^{a}$ and $E^{\alpha}{}_{|\alpha} = 0$;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S,\bar{g}_{\alpha\beta})$ is a space of constant curvature with scalar curvature $\bar{R} = -3k^{-1}$; and

(iii) $\Omega_{\alpha\beta} = 0$.

At this time we note that there is only a slight difference between the sets of characteristic initial data, I, II, III and IV, V, VI, resp. This similarity stems from the resemblance between the "magnetic constraint equations" (3.7), (3.10) and the "electric constraint equations" (3.9), (3.11).

Thus far in our search for characteristic initial data we have arranged for $\det Q$ to vanish by choosing $M^{\alpha\beta,\gamma} = 0$ and $\det(\chi^{AU}) = 0$. Due to Eq. (3.1) we see that we can also get $\det Q = 0$ by having $M^{\alpha\beta,\gamma} = 0$ and $\det(P^{\gamma\mu}) = 0$. We shall now explore this possibility.

When S is spacelike,

$$P^{\gamma\mu} = \bar{g}^{\gamma\mu} - k{}^{*}R^{*0\gamma 0\mu},$$

while if S is timelike,

$$P^{\gamma\mu} = -\bar{g}^{\gamma\mu} - k{}^{*}R^{*0\gamma 0\mu},$$

where ${}^{*}R^{*0\gamma 0\mu}$ is given by Eq. (3.8). Upon noting that $\det(P^{\gamma\mu}) = 0$ if and only if $\det(P^{\gamma}_{\mu}) = 0$, and

$$6\det(P^{\gamma}_{\mu}) = \delta^{\alpha\beta\gamma}_{\lambda\mu\nu}P^{\lambda}_{\alpha}P^{\mu}_{\beta}P^{\nu}_{\gamma},$$

we can derive the following two additional sets of characteristic initial data.

Set VII: Characteristic initial data when S is spacelike and $k < 0$.

(i) F_{ij} is pure magnetic;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S,\bar{g}_{\alpha\beta})$ admits B^{α} as a parallel vector field and has constant scalar curvature $\bar{R} = 3k^{-1}$; and

(iii) $\Omega_{\alpha\beta} = \Lambda\bar{g}_{\alpha\beta}$, where $\Lambda^{2} = -(2k)^{-1}$.

Set VIII. Characteristic initial data when S is timelike and $k > 0$.

(i) F_{ij} is pure magnetic;

(ii) $\bar{g}_{\alpha\beta}$ is such that $V_3 = (S,\bar{g}_{\alpha\beta})$ admits B^{α} as a parallel vector field and has constant scalar curvature $\bar{R} = 3k^{-1}$; and

(iii) $\Omega_{\alpha\beta} = \Lambda\bar{g}_{\alpha\beta}$, where $\Lambda^{2} = (2k)^{-1}$.

In concluding this section I would like to point out that the characteristic initial data presented above by no means exhausts the set of all possible characteristic initial data for the source-free generalized Einstein–Maxwell field equations. However, it does show us that characteristic initial data for these field equations can be prescribed on spacelike and timelike hypersurfaces.

4. SUMMARY AND CONCLUSIONS

In this paper we have seen that null hypersurfaces can act as characteristic surfaces for the source-free generalized Einstein–Maxwell field equations. We have also seen that it is possible to specify characteristic initial data on spacelike and timelike hypersurfaces. The latter observation does *not* imply that spacelike and timelike hypersurfaces can act as characteristic surfaces, although it is indicative of this possibility.

Previous work[11] with the generalized Einstein–Maxwell field equations suggests that the coupling constant k, which has units of (length)^2, must be small in magnitude. If we assume that this is the case, then we find that for each of the eight sets of characteristic initial data presented in the last section some quantity of physical or geometrical significance must be quite large. For example, we must have either intense electromagnetic fields, or enormous curvature or a rapidly varying metric. The fact that some quantity appears to be "blowing up" in order to obtain characteristic initial data on a spacelike or timelike hypersurface indicates that characteristic surfaces may only occur under extreme condi-

tions, during which a classical field theory may not actually be viable. Since our present knowledge of such situations is quite limited, I believe that we *cannot* dismiss the generalized Einstein–Maxwell field theory on the grounds that it admits characteristic initial data on spacelike or timelike hypersurfaces—unless we can produce more "reasonable" characteristic data on such hypersurfaces.

In Sec. 2 we saw that for noncharacteristic, analytic initial data, which satisfy the constraint equation (2.5) on a spacelike or timelike initial hypersurface, there will always exist an analytic solution to the source-free generalized Einstein–Maxwell field equations which is valid on a neighborhood of the initial hypersurface and which agrees with the initial data. When the initial data are characteristic, we are not assured that such a solution will exist, even if the data are analytic and satisfy the constraint equation. Thus in view of the characteristic initial data presented in Sec. 3, and the above discussion of it, we see that for intense gravitational and electromagnetic fields the source-free generalized Einstein–Maxwell field equations appear to lose their predictive powers. Consequently, these field equations seem to forecast the regime in which they are no longer useful for predicting the future. This remarkable feature of the generalized Einstein–Maxwell field theory can perhaps be interpreted as a classical limit on the phenomenon to which this theory can be applied, or perhaps it is indicative of something more fundamental about the nature of matter.

ACKNOWLEDGMENTS

I would like to thank J. Wainwright, D.A. Szafron, and S.J. Aldersley for many interesting discussions on the topics dealt with in this paper. I also wish to thank the Natural Sciences and Engineering Research Council of Canada for financial support.

[1]My notational conventions are the same as those employed in C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973) with the following exceptions: (i) tensorial indices will be denoted by lower case Latin letters, (ii) lower case Greek letters will assume the values 1–3 and satisfy the summation convention, and (iii) the permutation symbols will be denoted by ϵ^{abcd} and ϵ_{abcd}, with $\epsilon^{0123} = \epsilon_{0123} = 1$.

[2]For a discussion of the origin and uniqueness of the generalized Einstein–Maxwell field equations see: G.W. Horndeski, J. Math. Phys. **17**, 1980 (1976); and G.W. Horndeski and J. Wainwright, Phys. Rev. D **16**, 1691 (1977).

[3]See, e.g., p. 360–63 in J.L. Synge, *Relativity, The General Theory* (North-Holland, Amsterdam, 1971).

[4]For a more detailed discussion of the metric (1.4) see pp. 957–60 of Ref. 1.

[5]This solution represents a special case of the solution described in Footnote 20 of G.W. Horndeski, "Null Electromagnetic Fields in the Generalized Einstein–Maxwell Field Theory," J. Math. Phys. **20**, 726 (1979).

[6]The proof of this fact is virtually identical to the proof of the corresponding result in the Einstein–Maxwell field theory (see pp. 361–62 of Ref. 3) and does not require x to be a Gaussian normal coordinate system.

[7]An indirect proof of this claim can be given using Eq. (2.1) of the first reference cited in Ref. 2.

[8]The form of the Cauchy–Kowalewsky theorem which we need can be obtained by generalizing the formulation of this theorem presented on pp. 76–86 of F. John, *Partial Differential Equations* (Springer, New York, 1971).

[9]Recall that we always require our initial data to be such that $g_{00} = \pm 1$, $g_{0\alpha} = 0$, and $g_{0\alpha,0} = 0$. Thus, in order to specify the remaining initial data, all we need present is F_{ij}, $\bar{g}_{\alpha\beta}$ and $\Omega_{\alpha\beta}$.

[10]An n-dimensional pseudo-Riemannian space is said to be a space of constant curvature if $R_{abcd} = [R/n(n-1)](g_{ac}g_{bd} - g_{ad}g_{bc})$, where R is constant.

[11]See Sec. 4 of the second paper cited in Ref. 2.

Local limit theorem for Gibbs random fields of particles and unbounded spins

Massimo Campanino
Istituto Matematico G. Castelnuovo, Universitá di Roma, Roma, Italy

Gabriella Del Grosso
Istituto Matematico, Universitá di Camerino, Camerino, Italy

Brunello Tirozzi
Istituto Matematico, Universitá dell'Aquila, l'Aquila, Italy

In this paper we extend the results of 2 to Gibbsian systems of particles and unbounded spins, showing that under general conditions on the potential the local limit property is a consequence of the integral property for these systems. We give some applications of this result to the problem of the equivalence of the ensembles.

INTRODUCTION

In this work we find that under general conditions on the potential it is possible to derive the local limit theorem from the integral limit theorem for Gibbs random fields of particles and for Gibbs random fields of infinite spins.

We consider a system of particles interacting through a short-range pair potential; the proof consists in reducing the problem to the finite spin case using a special technique.[1] The proof of the local limit theorem for a Gibbs random field of particles has been derived directly by Halfina[2] and by Minlos and Halfina,[3] who have used analytic properties of the energy and of the correlation functions. Our result is quite different: Without assuming any analyticity we show that the local central limit theorem is a direct consequence of the integral one.

The case of infinite spins with a short-range pair interaction is easily reduced to the proof for the particles, using a property of exponential weak dependence introduced by Dobrushin,[4] explicitly exhibited in Ref. 5. Since the exponential weak dependence implies the exponential decay of correlations (see Appendix A) and Malyshev has shown that the exponential decay implies the integral limit theorem,[6] this result allows us to get sufficient conditions for the local limit theorem for infinite spin systems. The result of Riauba[7] is more general than ours in the one-dimensional case, whereas our simpler method works in any dimension.

From the local limit theorem we deduce the equivalence of the Gibbs ensemble and the canonical ensemble for spin systems and for particle systems. Our method differs from that used by Halfina[2] because the use of the exponential weak dependence makes the proof much simpler. Our work also differs from the result of Georgii,[8] because we find a method for checking the equivalence of the canonical ensemble in concrete cases. The result of Georgii is more general than ours, but since there is no method available for checking that a canonical state is extremal, it is more difficult to apply it to concrete models.

In Sec. 1, we give the definitions and the results. In Secs. 2 and 3 we give the proof of Theorems 1,2, and 3. In Sec. 4 we deal with the problem of the equivalence of ensembles for the systems we have considered in the previous sections. In Appendix A, as stated before, we show that the exponential decay of correlations follows from the exponential weak dependence for superstable infinite spin systems.

1. DEFINITIONS, HYPOTHESES, AND RESULTS

A. Particle system

Here we use the standard definitions of a Gibbsian system of particles (see Ref. 9). We assume a pair interaction, which is described by a translationally invariant potential $U:\mathbb{R}^{\nu}\to(-\infty,+\infty]$ which satisfies the following requirements:

(i) U is a continuous map into the extended real line;

(ii) For some $R>0$, $U(x)=0$ if $|x|>R$ (finite range) and for some

$$G>0 \text{ and every } x\in\mathbb{R}^{\nu},\ U(x)>-G; \tag{1.1}$$

(iii) $U(x-y)$ is stable, i.e., for some $B\geqslant 0$ and every finite configuration s,

$$\sum_{\substack{x,y\in s\\ x\neq y}} U(x-y)\geqslant -B|s|, \tag{1.2}$$

where $|s|$ denotes the cardinality of s. Let μ be the chemical potential and s, $\bar{s}$ two configurations; we get

$$U(s|\bar{s})=\tfrac{1}{2}\sum_{\substack{x,\bar{x}\in s\\ x\neq\bar{x}}} U(x-\bar{x})-\mu|s|+\sum_{\substack{x\in s\\ \bar{x}\in\bar{s}}} U(x-\bar{x}). \tag{1.3}$$

The conditional Gibbs distribution in a bounded volume V with the condition $s\in S(\mathbb{R}^{\nu}\setminus V)$ is given by the density with respect to the free measure $\lambda(ds)$[9]:

$$\begin{aligned} &P_V(s|\bar{s})=Z(V|\bar{s})^{-1}\exp\{-\beta U(s|\bar{s})\},\\ &Z(V|\bar{s})=\int_{S(V)}\exp\{-\beta U(s|\bar{s})\}\lambda(ds), \end{aligned} \tag{1.4}$$

where $S(V)d$ is the disjoint union of all the sets of configurations with a fixed number of particles:
$S(V)=U_{m=0}^{\infty}S^m(V)$, $S^m(V)\equiv\{s|S,|s|=m\}$, and S is the

 0022-2488/79/081752-07$01.00

family of all finite or countable subsets $s\in\mathbb{R}^\nu$ such that the intersection $s\cap V$ is finite for every bounded subset V of $\mathbb{R}^\nu$. The Gibbs random field associated to the particle system is a random field whose conditional distributions have densities (1.4) for almost all the conditions $\bar{s}$.

B. Infinite spin system (discrete values)

We first consider the case when the random variables ξ_t, $t\in Z^\nu$ take integer values.

The interaction is defined by a finite range potential (two-body) with range R, i.e., by a family of functions $U_{t,s}(\cdot,\cdot): Z\times Z\to(-\infty,+\infty]$ and by the self interaction potential $F_t: Z\to(-\infty,+\infty]$.

We suppose:

(i) $U_{t,s}, \forall t,s, F_t$ are translational invariant, (1.5)

(ii) if $|\xi|\leqslant\bar{c}\delta(2R+1)^\nu$, $|\eta|\leqslant\bar{c}\delta(2R+1)^\nu$, $\delta>1$, then $U_{t,s}(\xi,\mu)\leqslant R_1$, $|F(\xi)|\leqslant R_2$,

where $\bar{c}$ is defined in Eq. (1.10).

The Gibbs random field for an infinite spin system is defined as above,[4,10] and the conditional probabilities in the volume V with boundary conditions $\bar{\xi}_t$, $t\in Z^\nu\setminus V$ are

$$q_V(\xi_t, t\in V\,|\,\bar{\xi}_t, t\in Z^\nu\setminus V) = Z_V^{-1}(\bar{\xi}_t)\cdot\exp\Big\{+\beta\mu\sum_{t\in V}\xi_t - F(\xi_t) - \beta\sum_{\substack{t,s\in V\\ t\neq s}} V_{t,s}(\xi_t,\xi_s) - \beta\sum_{\substack{t\in V\\ s\in Z^\nu\setminus V}} U_{t,s}(\xi_t,\bar{\xi}_s)\Big\}. \tag{1.6}$$

C. Infinite spin system (continuous values)

Let us now give the conditions on the potential that we need to formulate our result. The symbols and the definitions of Sec. 1B will be used, taking into account that the conditional probability (1.6) now becomes a density of conditional probability with respect to the Lebesgue measure $\Pi_{l=1}^{|\nu|} d\xi_l$ and that all the variables in that formula take values in $\mathbb{R}^1$. Let us introduce the following symbols and definitions:

$$\begin{aligned} &P_{\bar{\gamma}}(\bar{x}) = P_{\{s\}}(x/\bar{\gamma}_l l\in A_s),\\ &A_s = \{l\in Z^\nu/|l-s|\in R\},\quad \bar{\gamma} = (\gamma_l, l\in A_s)\\ &C_\mu = \max_{\substack{|\bar{\gamma}_l|\leqslant\mu\\ l\in A_s}}\int_{-\infty}^{+\infty}|P^{\bar{\gamma}}_{\bar{\gamma}}(x)|dx, \end{aligned} \tag{1.7}$$

where $P_{\bar{\gamma}}(x) = P_{\{s\}}(x|\bar{\gamma}_l, l\in A_s)$ is the density of the conditional probability of the variable under the condition $\xi_l = \bar{\gamma}_l$, $l\in A_s$.

We suppose further that $U_{t,s}(\xi_t,\xi_s)$, $F(\xi_t)$ satisfy all the conditions of Sec. 1B and that they are differentiable with respect to ξ_t,ξ_s in such a way that $\int_{-\infty}^{+\infty}|P^{\bar{\gamma}}_{\bar{\gamma}}(x)|dx$ exists for every choice of the conditions $\bar{\gamma}_l, l\in A_s$ and that there exist constants $\alpha<2, g>0$ such that $C_\mu\leqslant g\mu^\alpha$. It is easy to see that these last two conditions hold, for example with $\alpha=1$, when the interaction is given by

$$U_{t,s}(\xi_t,\xi_s) = J(t-s)\xi_t\xi_s,\quad J(t)=0 \quad \text{if} \quad |t|\geqslant R,$$
$$F(\xi_t) = \lambda(a\xi^4 + b\xi^3 + c\xi^2) \tag{1.7a}$$

with $\alpha>0$.

D. The exponential weak dependence

Following Refs. 4 and 11 we introduce the property of exponential weak dependence for a Gibbs random fields associated to a spin system; an analogous definition holds for the particle case. We make use of the following metric on the space of measures (Vasershtein's distance). Let ξ and μ be two integer-valued random variables with distributions $P_\xi(\cdot), P_\mu(\cdot)$, respectively: We define the Vasershtein's distance $R(P_\xi(\cdot), P_\mu(\cdot))$ between P_ξ, P_μ by

$$R(P_\xi, P_\mu) = \inf_{\widehat{P}}\sum_{l,m}|l-m|\widehat{P}(l,m), \tag{1.8}$$

where the inf is taken over all the joint probability distributions $\widehat{P}(\xi,\mu)$ such that $\widehat{P}(\xi=n,\mu\in Z) = P_\xi(\xi=n)$, $\widehat{P}(\xi\in Z, \mu=m) = P_\mu(\mu=m)$. This distance has been used by Dobrushin[4] in a more general frame than ours for studying the problem of the uniqueness of the Gibbs random field, also see Ref. 5.

Let V_1, $V_2\in Z^\nu$, $|V_1|<+\infty$, $|V_2|<+\infty$ be such that $V_1\cap V_2=\emptyset$ and take the constant $\epsilon>0$. We shall say that the values of the field on V_2 are exponentially weakly dependent of the values of the field on V_1 if there exist numbers $\epsilon_t, t\in V_2$, s.t. $\Sigma_{t\in V_2}\epsilon_t\leqslant\epsilon$ and for every $\bar{V}_1\subset V_1$ and P_{V_2} every $x_t^1, x_t^2\in Z$, $t\in V_2$

$$R(P_{\bar{V}_1}|x_t^1, t\in V_2, P_{\bar{V}_1}|x_t^2 t\in V_2)\leqslant\sum_{t\in V_2}\epsilon_t|x_t^1 - x_t^2| \tag{1.9}$$

and $\epsilon_t\leqslant c_1\exp[-c_2 d(t,V_1)]$ for some fixed constants c_1, c_2, and $d(t,V_1)$ and $d(t,V) = \min_{t'\in V} d(t,t')$.

E. The integral and the local central limit theorems

Here we use the same notations of Ref. 11. In all the theorems that we prove in Secs. 2 and 3, we consider as a Gibbsian sequence the sequence of the restrictions of the Gibbs field to the space of the events related to the volume V_k, and the Gibbs field chosen is supposed to be invariant under translations. All the proofs also hold in the case when we consider the sequence of conditional probabilities $P_k(\cdot/\bar{x}_k)$ of the spins in the volume V_K provided that one chooses the boundary conditions $\bar{x}_k$ in such a way that for some $\bar{c}$, the expectation of $\Sigma_{t\in V_k}|\xi_t|$ with respect to $P_k(\cdot\bar{x}_k)$ satisfies the inequality

$$E_k\Big(\sum_{t\in V_k}|\xi_t|\Big)\leqslant\bar{c}|V_k|. \tag{1.10}$$

Thus $\bar{c}$ (the constant introduced in Sec. 1B) is the constant for which (1.10) is verified, or, in the case of the translationally invariant Gibbs random field, ξ_t is defined by $\bar{c} = E|\xi_0|$. In the case of particles (1.10) must be substituted by

$$E(S_k)\leqslant\bar{c}m(V_k), \tag{1.10'}$$

where $S_k = S_n \cap V_k$. Let $\{S_k\}$ be a sequence of random variables such that S_k is measurable with respect to the σ-algebra of the events in the volume V_k. We shall say that this sequence satisfies the integral central limit theorem if

$$(\alpha)\ D(S_k) = D|V_k|,$$

$$(\beta)\ D > 0,$$

$$(\gamma)\ \Pr\{\bar{S}_k \leqslant x\} \underset{k\to\infty}{\to} \frac{1}{\sqrt{2\pi}}\int_0^x e^{-(u^2/2)}du,$$

where D denotes the dispersion

$$D\xi = \int[\xi(\omega) - E\xi(\omega)]^2 dP(\omega)$$

and

$$\bar{S}_k = (S_k - ES_k)/\sqrt{DS_k}.$$

In the particle case one has to substitute $|V_k|$ with V (volume of V_k).

In the case of lattice distributed random variables,[7,12] the local limit theorem is defined in the following way: Setting $P_k(p) = P_r\{S_k = p\}$, $z_p^k = (p - ES_k)/(DS_k)^{1/2}$ we say that the Gibbsian sequence $\{S_k\}$ satisfies the local central limit theorem if (α), (β) are verified and

$$\sup_p \left| \sqrt{DS_k}\, P_k(p) - \frac{1}{\sqrt{2n}} \exp[-\tfrac{1}{2}(z_p^k)^2] \right| \underset{k\to\infty}{\to} 0. \tag{1.11}$$

In the particle case we shall take $S_k = |s \cap V_k|$, in the infinite spin case $S_k = \Sigma_{t\in V_k}\xi_t$. When the spin is continuous it is necessary to introduce the density of conditional probability and the density of probability of S_k:$\Pr\{S_k \in A\} = \int_A P_k(x)dx$, where A is a Borel subset of the real numbers and dx the Lebesgue measure. Thus the condition (1.11) becomes

$$\sup_x \left| P_k(x) - \frac{e^{-(x^2/2)}}{\sqrt{2\pi}} \right| \underset{k\to\infty}{\to} 0. \tag{1.12}$$

F. Equivalence of the ensembles

We use definitions analogous to those used in Ref. 11. In this part of the work we deal with systems of spins which take values in $\mathbb{Z}$. The other systems can be treated easily using the same methods developed here.

Gibbs ensemble: In the volume V_k, the probability measure $P_{\mu,k}$ belonging to a certain Gibbsian sequence, where μ indicates the chemical potential, is called the Gibbs ensemble. We indicate with $P_{\mu,k}(\xi_t, t\in V)$ the restriction of $P_{\mu,k}$ on the space $\mathbb{Z}^u$, $u\in V_k$. We denote the mean and the variance of the random variable X with respect to the measure $P_{\mu,k}$ by $E_{\mu,k}X$, $D_{\mu,k}X$.

Canonical ensemble: We shall say that the probability measure on Z^{V_k} denoted by $q_{N,V_k}(\cdot)$ is the canonical ensemble if $q_{N,V_k}(\cdot)$ is given by

$$q_{N,V_k}(\xi_t, t\in V = P_{\mu,k}(\xi_t, t\in V / S_k = N), \tag{1.13}$$

where $S_k = \Sigma_{t\in V_k}\xi_t$ and $P_{\mu k}(\cdot|S_k = N)$ is the conditional distribution obtained by $P_{\mu k}$ with the condition $S_k = N$, and N is some integer such that $P_{\mu,k}(S_k = N) \neq 0$.

We call a Gibbsian canonical sequence a sequence of canonical ensembles generated by a Gibbsian sequence and by a sequence of integers.

Equivalence between Gibbs ensemble and canonical ensemble: Let us take a Gibbsian sequence $\{S_k\}$. Let N_K be a sequence of integers such that $\lim_{k\to\infty} N_k/|V_k| = \rho$. ($\rho$ is the density in the canonical ensemble.) We say the equivalence between the Gibbs ensemble and the canonical ensemble holds, for the given Gibbsian sequence, if, for every U, $|U| < +\infty$ and for every ξ_t

$$|P_{\mu_k}(\xi_t, t\in U) - q_{N_k,V_k}(\xi_t, t\in U)| \underset{k\to\infty}{\to} 0, \tag{1.14}$$

where μ_K is such that $E_{\mu_k}S_k = N_k$.

G. Results

Using the previous definitions we now formulate the theorems contained in this paper.

Theorem 1: Let the hypotheses of (1.1) on the potential be verified. If $\{S_k\}$ verifies the integral central limit theorem for a certain Gibbsian sequence, then it also verifies the local limit theorem for the same Gibbsian sequence.

The case of infinite spins with discrete values can be proved in an analogous way, when the hypothesis of the following lemma are verified.

Lemma 1: If the property of exponential weak dependence is verified by the Gibbs random field defined in (1.2), then for some $q > 0$

$$D\sum_{t\in V_k} |\xi_t| \leqslant q|V_k|. \tag{1.15}$$

We need this additional hypothesis in the case of infinite spins because we have to evaluate the probability that a certain fraction of the random variables in a volume V_k has absolute value bigger than a fixed constant. In the particle case (1.15) is satisfied by condition (α) of the integral central limit theorem. Thus we obtain the following theorem for infinite discrete spins:

Theorem 2: If the exponential weak dependence is valid for the Gibbs random field $\{\xi_t\}$ satisfying the hypotheses in (1.2) and the sequence $\{S_k\}$ satisfies the integral central limit theorem for a certain Gibbsian sequence, then it satisfies the local central limit theorem for the same Gibbsian sequence.

In the case of the Gibbs random field of particles, it is possible to derive (α) and (γ) defined in (1.4) using the analyticity property of the partition functions as a function of the chemical potential, while (β) can be deduced from the strong convexity of the pressure. Thus we introduce the integral and the local limit theorem in the analyticity region for a Gibbs random field of particles. Let us formulate here the result concerning the connection between exponential weak dependence and the exponential decay of correlations.

Lemma 2: Let the Gibbs random field (Sec. 1B) satisfy the exponential weak dependence and suppose that the superstability property is satisfied.[10,13] Then for every $T_1, T_2 \in \mathbb{Z}^\nu, |T_1| < +\infty, |T_2| < +\infty, T_1 \cap T_2 = \emptyset$ we have

$$\left|E\left(\prod_{t\in T_1\cup T_2}\xi_t\right)-E\left(\prod_{t\in T_1}\xi_t\right)E\left(\prod_{\xi\in T_2}\xi_t\right)\right|\leqslant b_1 e^{-b_2 d(T_1,T_2)}$$

where the constants e_1, e_2 depend only on $|T_1\cup T_2|$ and

$$d(T_1,T_2)=\min_{t\in T_1,t_2\in T_2}\operatorname{dist}(t_1,t_2).$$

The proof of Lemma 2 can be extended also to the spin system with continuous values. Also theorem 2 can be extended to this case.

Theorem 3: Let us consider a Gibbs random field of spins satisfying the hypothesis in Sec. 1C. Then if the sequence $\{S_k\}$ satisfies the integral central limit theorem for a Gibbsian sequence and the property of exponential weak dependence is verified, then it satisfies also the local central limit theorem for the same Gibbsian sequence.

Theorems 2 and 3 together with Lemma 2 furnish sufficient conditions in order to obtain the integral and the local limit theorem. Let us consider, for example, the Gibbsian sequence defined by the restriction of the Gibbs random field (Sec. 1C) to a sequence of volumes $\{V_k\}$. Let us take the Gibbs field corresponding to potential (1.7a) which satisfies the superstability property. Then it can be shown[5] that the exponential weak dependence is satisfied at large temperatures and applying Lemma 2, we obtain the exponential decay of correlations. From the last property, applying the methods used in Ref. 6, we obtain the integral central limit theorem for the choosen Gibbsian sequence. Then from Theorem 3 the local theorem is proven. From these three theorems we can derive the equivalence between the Gibbs ensemble and the canonical ensemble. We consider only the case of spin systems with integer values, the other cases can be treated in an analogous way. We first take a Gibbsian sequence $P_{\mu,k}$ and then construct the canonical Gibbsian sequence (G_{N_k,V_k},N_k) as in Sec. 1F. Then we have:

Theorem 4: If the Gibbsian sequence $\{P_{\mu,k}\}$ satisfies (1.10) and the integral central limit theorem, if the potential satisfies conditions (1.2), if the exponential weak dependence holds, then the equivalence between canonical and Gibbs ensemble holds for $\{P_{\mu,k}\}$. First we prove the proposition:

Proposition 1: If for the Gibbsian sequence $\{P_{\mu,k}\}$, the integral central limit theorem holds, and if the exponential weak dependence is verified, then $\forall\, Q$, $|Q|<+\infty$, $\forall\, \xi_t$, $t\in Q$,

$$P\left(\frac{S_{V_k}|-E_{\mu,k}S_{V_k}}{\sqrt{D|V_k|}}<z\right)|\xi_t t\in Q\to\int_{-\infty}^{z}\frac{e^{-u^2/2}}{\sqrt{2\pi}}\,du, \tag{1.16}$$

where D is defined in (α), Sec. 1E. Theorem 4 follows from Proposition 1 and Theorem 2 by standard arguments.

2. PROOF OF THEOREM 1

The probability $P_k(p)=P_r\{S_k=p\}$ can be expressed by means of the inversion formula for the characteristic function

$$2\pi P_k(p)=\frac{1}{\sqrt{DS_k}}\int_{-\pi\sqrt{DS_k}}^{+\pi\sqrt{DS_k}}E\left(e^{i\tau\bar{S}_k}\right)e^{-i\tau p}\tau. \tag{2.1}$$

We use the same subdivision of Ref. 11. Therefore, the problem reduces to showing the following: Given $\epsilon>0$ there are two constants A and χ, such that for k large enough the integrals:

$$\int_{A<|t|<\chi\sqrt{DS_k}}|E(\exp(i\tau\bar{S}_k))|d\tau, \tag{2.2a}$$

$$\int_{\chi\sqrt{DS_k}\leqslant|t|\leqslant\pi\sqrt{DS_k}}|E(\exp(i\tau\bar{S}_k))|d\tau, \tag{2.2b}$$

are less than ϵ.

Let us consider (2.2a). It follows from the continuity of U that for some $R_2>0$, $R_2<R_1$, $U(x)<G$ for $x>R_2$. Let's now put $R=\max(R_1,3R_2)$. We consider a pavement $\mathscr{P}$ of $\mathbb{R}^\nu$ made up of open cubes with side $3R$, a subpavement $\mathscr{Q}$ of $\mathscr{P}$ made up of open cubes with side R, and let's denote with $\mathscr{Q}_1$ the set of all cubes of $\mathscr{Q}$ that lie at the center of a cube of $\mathscr{P}$.

We put

$$W_k=\bigcup_{\substack{P\in\mathscr{P}\\ P\subset V_k}}P,\quad Z_k=V_k\setminus W_k,$$
$$n_k=\#(P\in\mathscr{P},\,P\subset V_k). \tag{2.3}$$

The event that some particle lies on the boundaries of the cubes has probability 0 with respect to every Gibbs random measure, so that we will neglect it in the following considerations.

We have

$$S_k=S_{W_k}+S_{Z_k}. \tag{2.4}$$

Now since S_k satisfies the integral theorem and in particular the property α, it follows from Chebychev's inequality and (1.10)′ that for an arbitrary $\eta>0$

$$P\left(\frac{\Sigma_{\Delta\subset W_k}S_\Delta}{3^\nu n_k}\geqslant\bar{c}R^\nu+\eta\right)\leqslant P\left(\frac{S_k}{3^\nu n_k}\geqslant\bar{c}R^\nu+\eta\right)\leqslant\frac{Q}{n_k}, \tag{2.5}$$

where Q is some positive constant.

Let us fix two constants δ, δ' such that $\delta>3^\nu$ and $3^{-\nu}>\delta'>\delta^{-1}$ and define $M=\delta(\bar{C}R^\nu+\eta)$ and χ_μ the indicator of the half-plane $[M,+\infty)$. It is easy to verify that the event

$$\mathscr{A}_k=\left\{\frac{1}{3^\nu n_k}\sum_{\substack{\Delta\in\mathscr{Q}\\ \Delta\subset W_k}}\chi_M(S_\Delta)\geqslant\delta'\right\} \tag{2.6}$$

is contained in the event

$$\mathscr{B}_k=\left\{\frac{1}{3^\nu n_k}\sum_{\substack{\Delta\in\mathscr{Q}\\ \Delta\subset W_k}}S_\Delta\geqslant\bar{C}R^\nu+\eta\right\}, \tag{2.7}$$

so that it follows from (2.5) that

$$P(\mathscr{A}_k)\leqslant\frac{Q}{n_k}. \tag{2.8}$$

The event $\mathscr{B}_k$ can be decomposed as the union of a family $\{\sigma_F\}$ of mutually disjoint events, where F ranges over the

sets of cubes of $\mathcal{Q}$ contained in W_k such that $\#(F)\leqslant\delta' 3^v n_k$ and

$$\sigma_F = \bigcap_{\Delta\in F}(S_\Delta \geqslant M) \bigcap_{\Delta\subset W_k\setminus F}(S_\Delta < M). \tag{2.9}$$

We can write

$$|E(\exp(i\tau\overline{S}_k))| \leqslant |E(\exp(i\tau\overline{S}_k|\mathcal{A}_k))|P(\mathcal{A}_k) + \sum_{\substack{F\subset W_k \\ \#(F)<\delta' 3^v n_k}} |E(e^{i\tau\overline{S}_k}|\sigma_F)|P(\sigma_F). \tag{2.10}$$

Let us condition the variables S_Δ $\Delta\in\mathcal{Q}_1\cap W_k$ with a configuration $S\subset V_k\setminus\mathcal{Q}_1$ and some event σ_F. The finite range of the potential implies that these conditioned variables are independent and the characteristic function of their sum factorizes. We also observe that if $\#(F)<\delta' 3^v n_k$, there are at least $n_k(1-\delta' 3^v)$ cubes Δ in $\mathcal{Q}_1\cap F$ such that the neighboring cubes of Δ also belong to F.

We have

$$|E(e^{i\tau\overline{S}_k}|\sigma_F)| = \left|E\left(E\left\{\exp[i\tau]\left(\sum_{\Delta\subset W_k\cap\mathcal{Q}_1}\frac{S_\Delta}{\sqrt{DS_k}}\right)\times\exp\left(i\tau\frac{S_{V_k\setminus\mathcal{Q}_1}}{\sqrt{DS_k}}\right)\middle|\sigma_F, S\right\}\middle|\sigma_F\right)\right|. \tag{2.11}$$

We can majorize (2.11) by taking the modulus inside the first expectation and, since the second factor in the expression (2.11) is a function of the conditions, we obtain that (2.11) is less than or equal to

$$E\left(\prod_{\Delta\subset W_k\cap\mathcal{Q}_1}\left|E\left[\exp\left(i\tau\frac{S_\Delta}{\sqrt{DS_k}}\right)\middle|\sigma_F, S\cap\left(\bigcup_{\substack{\Delta'\in\mathcal{Q}/\mathcal{Q}_1 \\ d(\Delta,\Delta')=0}}\Delta'\right)\middle|\sigma_F\right]\right|\right) \tag{2.12}$$

and (2.12) is bounded by

$$\left\{\sup_{\substack{s\subset\cup\Delta' \\ d(\Delta',\Delta)=0 \\ \Delta'\in\mathcal{Q}\setminus\mathcal{Q}_1 \\ |\Delta|<3^v M}}\left|E\left[\exp\left(i\tau\frac{S_\Delta}{\sqrt{DS_k}}\middle|s; S_\Delta\leqslant M\right)\right]\right|\right\}^{n_k(1-\delta' 3^v)}, \tag{2.13}$$

where Δ is any element of $\mathcal{Q}_1$.

We make a Taylor's expansion around 0:

$$\left|E\left(\exp\frac{i\tau}{\sqrt{DS_k}}(S_\Delta|s, S_\Delta\leqslant M)\right)\right| = 1 - \frac{t^2}{DS_k}\widetilde{DS}_\Delta + o\left(\frac{t^2}{DS_k}\right), \tag{2.14}$$

where $\widetilde{D}$ is the conditional variance.

It is easy to see that

$$\widetilde{DS}_\Delta \geqslant \tfrac{1}{2}\min[P(S_\Delta=0|s, S_\Delta\leqslant M), P(S_\Delta=1|s, S_\Delta\leqslant M)] \tag{2.15}$$

and a standard computation shows that our conditions on the potential give to the probabilities on the right-hand side of (2.15) the bounds:

$$P(S_\Delta=0|s,\ S_\Delta\leqslant M) \geqslant \exp[-R^v\exp(\beta_\mu+\beta B/2+3^v\beta GM)],$$

$$P(S_\Delta=1|s, S_\Delta\leqslant M) \geqslant (R/3)^v\exp(-\beta G 3^v M+\mu B - R^v\exp|\beta\mu+\beta B/2+3^v\beta GM), \tag{2.16}$$

if $|s|\leqslant 3^v M$.

Since the expectation (2.14) is computed with the condition $S_\Delta\leqslant M$, the conditional third moment is uniformly bounded and the quantity can be bounded by an infinitesimal term independent from the conditions.

From (2.14), (2.15), and (2.16) it follows that there are two numbers $\alpha>0$ and $\chi>0$ such that (2.14) is bounded by

$$\exp\left(-\tau^2\frac{\alpha}{DS_k}\right) \quad \text{for } |\tau|\leqslant\chi\sqrt{DS_k}. \tag{2.17}$$

Taking account of the estimates (2.12) and (2.17) and of the formula (2.13) we obtain that for a suitable constant $c>0$,

$$|Ee^{i\tau\overline{S}_k})| \leqslant \exp(-c\tau^2)+Q/n_k,$$

and for A and k large enough

$$\int_{A\leqslant|\tau|\leqslant\chi\sqrt{DS_k}}|E(e^{i\tau\overline{S}_k})|d\tau \leqslant \int_A^\infty \exp\{-c\tau^2\}d\tau + \pi\chi\sqrt{DS_k}\frac{Q}{n_k}\leqslant\epsilon.$$

The second integral of (2.2) can be bounded without difficulty by means of the same method and of the estimate of Ref. 11.

3. PROOFS OF THEOREMS 2 AND 3

In the case of infinite spins with discrete values we need only to prove Lemma 1, since the proof of Theorem 2 is completely analogous to that of Theorem 1.

A. Proof of Lemma 1

We have

$$\begin{aligned} DS_{V_k} &= E\left(\sum_{t\in V_k}(|\xi_t|-E|\xi_t|)\right)^2 \\ &\leqslant \sum_{t,s\in V_k}\left|\sum_{e,m}(|m|\,|l|P(\xi_t=m,\xi_s=l) - |m|\,|l|P(\xi_t=m)P(\xi_s=l))\right| \\ &\leqslant \sum_{t,s\in V}\sum_l |l|P(\xi_j=l)\left[\left|\sum_m |m|P(\xi_t=m/\xi_s=l) - mP(\xi_t=m)\right|\right]. \end{aligned} \tag{3.1}$$

Now

$$\begin{aligned} &\sum_m(|m|P(\xi_t=m/\xi_s=l)-|m|P(\xi_t=m)| \\ &\leqslant \sum_n P(\xi_s=n)\left|\sum_m |m|P(\xi_t=m/\xi_s=l) - \sum|m|P(\xi_t=m/\xi_s=n)\right|. \end{aligned} \tag{3.2}$$

and we can majorize the expression under the modulus with the Vasershtein distance using the inequality

$$\sum |l-m|\,\hat{P}(l,m) \geqslant \sum ||l|-|m||\hat{P}(l,m)$$

$$\geqslant \left|\sum_l |l|P_\xi(l) - \sum |m|P_\eta(m)\right|, \qquad (3.3)$$

valid for any two random variables with joint distribution $\hat{P}(l,m)$. Using the inequalities (3.1), (3.2), and (3.3) and condition (1.8) of Vasershtein's distance we obtain

$$DS_{V_k} \leqslant \sum_{t,s\in V_k}\sum_l |l|P(\xi_s = l)\sum_k P(\xi_s = k)$$
$$R\,(P_{|t|}(\cdot|\xi_s = l), P_{|t|}(\cdot|\xi_s = k)) \leqslant q|V_k|.$$

Since we are using the definition of the Gibbsian sequence given in Sec. 1E, Q.E.D.

We give now a sketch of the proof of Theorem 3.

Proof of Theorem 3: In the case of infinite continuous spins the integral which expresses the density $P_k(x)$ of S_k in terms of its characteristic function extends to the whole line,

$$P_k(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} E(e^{it\bar{S}_k})e^{-itx}dt. \qquad (3.4)$$

We use the same method of Theorem 1 and of Ref. 11 but we add a new integral to the subdivision, so that we have to estimate the following integrals:

$$\int_{\leqslant|\tau|\leqslant\chi\sqrt{DS_k}} |E(e^{i\tau\bar{S}_k})|d\tau, \quad \int_{\chi\sqrt{DS_k}\leqslant|\tau|\leqslant\rho\sqrt{DS_k}} |E(e^{i\tau\bar{S}_k})|d\tau,$$
$$\int_{\tau\geqslant\rho\sqrt{DS_k}} |E(e^{i\tau\bar{S}_k})|d\tau, \qquad (3.5)$$

where the constants A, χ, ρ can be suitably chosen. The first integral can be estimated in the same way as in Theorem 1. For the second integral we use the estimate

$$\left|E\left(e^{i\tau\frac{3i}{\sqrt{DS_k}}\bar{\gamma}_i}\right)\right| \leqslant e^{-\bar{c}}, \quad |\tau|\geqslant\chi\sqrt{DS_k} \qquad (3.6)$$

for the characteristic function of the central spin of a cube with conditions $\bar{\gamma}_i$, such that $|\bar{\gamma}_i|\leqslant M\ \forall\ i$ (see Ref. 14). We therefore obtain with the method of "good" and "bad" cubes the estimate

$$\int_{\chi\sqrt{DS_k}\leqslant|\tau|\leqslant\rho\sqrt{DS_k}} |E(e^{i\tau\bar{S}_k})|d\tau \leqslant (\rho-\chi)\sqrt{DS_k}$$
$$\times\left(\frac{Q}{V_k} + e^{-\bar{c}n_k(1-\delta')}\right). \qquad (3.7)$$

For what concerns the third integral we divide the domain of the integration into intervals of equal length $J_n = [\rho\sqrt{DS_k} + (n-1)\rho\sqrt{DS_k} + n]$. In each of this intervals we apply the method of "good" and "bad" cubes, but the events $\mathscr{A}_k^{(n)}$ vary with n,

$$\mathscr{A}_k^{(k)} \equiv \left\{\frac{1}{|V_k|}\sum X_i^{(n)} \geqslant \delta'\right\}$$

where

$$X_i^{(n)} = \begin{cases} 1, & \text{if } |\xi_i| > M_n, \\ 0, & \text{if } |\xi_i| \leqslant M_n \end{cases},$$

where the constant $M_n = \bar{c} + \eta_n$ will be chosen later. For the characteristic function of the central spin of a cube we use the estimate

$$|f(t)| \leqslant \frac{1}{t}\int_{-\infty}^{+\infty} |P'(x)|d|x$$

where $P'(x)$ is the density of a random variable and $f(t)$ is its characteristic function. Using the last remark and the hypothesis in Sec. 1G on the potential we obtain:

$$\int_{\rho\sqrt{DS_k}}^{+\infty} |E(e^{i\tau\bar{S}_k})|d\tau \leqslant \frac{1}{\sqrt{|V_k|}} D^{3/2}\sum_{n=1}^{\infty}\frac{1}{\eta_n^2}$$
$$+ \sqrt{DV_k}\sum_{n=1}^{\infty}\left(\frac{\beta\eta_n^\alpha}{\rho+n-1}\right)^{n_k(1-\delta')},$$

which, with the choice $\mu_n = (\rho+n-1)^{1/2+\lambda}$, with $0<\lambda<(1-\alpha/2)\alpha^{-1}$ and ρ sufficiently large, can be easily shown to be less of any ϵ for $k > \bar{k}(\epsilon)$. Q.E.D.

4. EQUIVALENCE OF ENSEMBLES

A. Proof of Proposition 1

Without losing generality we suppose than the length of the side of V_k is $2k$. We consider a sequence of cubes $\{D_k\}$ with equal centers and sides $2[(V_k)^{1/2}]$. It is easy to see then that it is sufficient to consider the asymptotic behavior of the conditional distribution of

$$\bar{S}_{W_k} = \frac{1}{\sqrt{D|V_k|}}\sum_{t\in W_k}(\xi_t - E_{\mu_k,k}\xi_t), \qquad (4.1)$$

where $W_k = V_k|D_k$, since it is equal to that of $\bar{S}_{V_k}$. Now the distance of the set Q from W_k tends to ∞, as k goes to ∞ and it is straightforward to see that the Vasershtein distance between the unconditional and the conditional distribution of $\bar{\xi}_{W_k}$ is majorized for k big enough by the Vasershtein distance between the corresponding distributions of the set of the $\xi_t, t\in W_K$. Since the convergence w.r.t. Vasershtein distance implies weak convergence, (1.16) is true.

APPENDIX A

Proof of the Lemma 2: Let us consider the expression

$$\left|E\left(\prod_{t\in T_1\cup T_2}\xi_t\right) - E\left(\prod_{t\in T_1}\xi_t\right)E\left(\prod_{t\in T_2}\xi_t\right)\right|. \qquad (A1)$$

we can majorize it by

$$\left|\int_{t\in T_1\cup T_2}\prod\xi_t\chi_\mu(\xi_t)dP - \left(\int\prod_{t\in T_1}\xi_t\chi_\mu(\xi_t)dP\right)\right.$$
$$\left.\times\left(\int\prod_{t\in T_2}\xi_t\chi_\mu(\xi_t)dP\right)\right|_t + \sum_{T\neq\emptyset}\left\{\int[\xi_t\chi_\mu(\xi_t)]\right.$$
$$\left.\times\left[\prod_{t\in T_1\cup T_2\setminus T}\xi_t(1-\chi_\mu(\xi_t))\right]dP\right| + \left[\int\prod_{t\in T_1\cap T}\xi_t\chi_\mu(\xi_t)\right.$$
$$\left.\times\prod_{t\in T_1\setminus T}\xi_t(1-\chi_\mu(\xi_t))dP\right]\left[\int\prod_{t\in T_2\cap T_2}\xi_t\chi_\mu(\xi_t)\right]$$
$$\left.\times\left[\prod_{t\in T_2\setminus T}\xi_t\,\chi_\mu(\xi_t)\right]dP\right\}, \qquad (A2)$$

where M is a constant that will be suitably chosen later and $\chi_\mu(x)$ is the indicator of the interval $[-M,M]$. The first

term in (A2) is proportional to

$$\left| \int dP(\xi_t, t\in T_2) dP(\mu_t, t\in T_2) \times \left[\prod_{t\in T_1} \xi_t \chi_\mu(\xi_t) dP(\xi_t, t\in T_1) | \xi_t, t\in T_2) - \prod_{t\in T_1} \chi_\mu(\xi'_t) dP(\xi'_t, t\in T_1 | \mu_t, t\in T_2) \right] \right|. \quad \text{(A3)}$$

We can evaluate this expression in terms of Vasershtein's distances by considering a joint probability distribution for ξ_t and ξ'_t for $t\in T$ with the prescribed marginal distributions $P(\xi_t t\in T_1/\xi_t, t\in T_2)$ and $P(\xi'_t, t\in T_2/\mu_t t\in T_2)$ and majoring the difference of the products with the difference of two factors times $M^{|T_1|-1}$. There are new terms arising when only one factor is greater than M, but these terms are easily evaluated by means of the superstability inequality

$$P(\xi_t \geqslant x) \leqslant e^{-x^2 \cdot a + b}. \quad \text{(A4)}$$

In this way we obtain for (A3) the following majorization:

$$F|T_2|M^{|T_2|+|T_1|}(C_1 e^{-c_1 d(T_1,T_2)}) + |T_1|M^{|T_1|+|T_2|} e^{-\mu^2 a + b}, \quad \text{(A5)}$$

where F is a constant.

For what concerns the remaining terms of (2.15), their number is $2^{|T_1 \cup T_2|} - 1$ and each of them can be easily bounded by means of (A4) and of the Schwartz's inequality. They can be bounded by

$$2(2^{|T_1 \cup T_2|} - 1) M^{T_1 \cup T_2} e^{-aM^2 + b}. \quad \text{(A6)}$$

Choosing $M = d(T_1, T_2)$ and using (A5) and (A6), we prove the desired result.

It is clear that the proof of Lemma 2 applies with the obvious changes to the case of continuous spins.

ACKNOWLEDGMENTS

One of us (B. Tirozzi) acknowlegdes useful discussions with Professor R.L. Dobrushin. We acknowledge also useful discussions with Professor E. Presutti and Professor Benfatto, and thank Professor G. Gallavotti for a critical reading of our first manuscript.

[1] R.L. Dobrushin and P.A. Minlos, Teor. Ver. Prim. **12**, 595 (1967).
[2] A.M. Halfina, Mat. Sborn. **80**, 122 (1969).
[3] R.A. Minlos and A.M. Halfina, Isv. Akad. Nauk U.S.S.R. **34**, 1173 (1970).
[4] R.L. Dobrushin, Teor. Ver. Prim. **3**, 15 (1970).
[5] M. Cassandro, E. Olivieri, S. Pellegrinotti, and E. Presutti, Preprint.
[6] V.A. Malyshev, Dokl. Akad. Nauk U.S.S.R. **224**, 32 (1975).
[7] B. Riauba, Lit. Mat. Sb. **207**, 1 (1963).
[8] H. Otto Georgii, Commun. Math. Phys. **48**, 31 (1976).
[9] R.L. Dobrushin, Teor. Mat. Fis. **4**, 1, 101 (1970).
[10] D. Ruelle, Commun. Math. Phys. **50**, 189 (1976).
[11] R.L. Dobrushin and B. Tirozzi, Commun. Math. Phys. **54**, 173 (1977).
[12] B.V. Gnedenko, *Theory of Probability* (MIR, Moscow, 1976), Eng. trans.
[13] J.L. Lebowitz and E. Presutti, Commun. Math. Phys. **50**, 195 (1976).
[14] I.A. Ibragimov and Iu.V. Linnik, Nesavisimi i Stazionarni sviasiani sluciani velicini (Nauka, Moscow, 1965).

Solution of renormalization group recursion relations for a finite Ising chain

A. Rabinovitch
Department of Physics, Ben Gurion University of the Negev, Beer Sheva 84120, Israel

R. N. Sen
Department of Mathematics, Ben Gurion University of the Negev, Beer Sheva 84120, Israel

(Received 6 February 1979)

The recursion relations obtained by applying the renormalization group technique to a finite one-dimensional Ising chain with nearest-neighbor interaction and constant external magnetic field are decoupled and solved in closed form.

1. INTRODUCTION

It is fairly evident that knowledge of a renormalization group transformation [1,2] for a *finite* spin system gives, in principle, a method for calculating the partition function of the system by recursion. The method is as follows. Write the partition function for a system of N spins as

$$Z = \sum U(A_0^{(1)},...,A_0^{(k)};N), \tag{1}$$

where U depends on the k constants $A_0^{(l)}$, $l = 1,...,k$, some of which may be zero. The summation is over all spins $S_i = \pm 1$, $i = 1,...,N$, $S_{N+1} = S_1$, which have not been explicitly indicated in (1). Assume, for definiteness, that the renormalization group (hereafter abbreviated RG) transformation consists of summing over every other spin. Applying it n times, we obtain

$$Z = \sum U(A_n^{(1)},...,A_n^{(k)};N/2^n), \tag{2}$$

where the functional form of U does not change at any step; only the constants $A_m^{(l)}$ do. If n is large enough, i.e., if $N/2^n$ is small enough, then the summation over the few remaining spins in (2) can be carried out directly. Thus, if one has explicit expressions for the constants $A_n^{(1)},...,A_n^{(k)}$ in terms of the quantities $A_0^{(1)},...,A_0^{(k)}$, the problem is solved.

Recursion relations for the $A_{m+1}^{(l)}$ in terms of the $A_m^{(l)}$, $l = 1,2,...,k$, are provided by the RG transformations. These relations are coupled and highly nonlinear, and there appear to be *no* general methods for solving such equations *in closed form*. It may therefore be that such equations are worth studying *per se*.

In the present note we shall report on the strictly technical problem of the exact solution of the recursion relations which arise in one-dimensional Ising models with nearest-neighbor interaction and constant external magnetic field. The resolution of these equations, which were obtained by Nelson and Fisher, [1] depends on the exploitation of their symmetries to decouple them.

2. THE SOLUTION PROCEDURE

The recursion relations under consideration (see Appendix A) are

$$F_{n+1} = (F_nL_n + 1)(F_n + L_n)F_n^{-1}(L_n + 1)^{-2}, \tag{3a}$$

$$L_{n+1} = L_n(F_nL_n + 1)(F_n + L_n)^{-1}, \tag{3b}$$

$$Q_{n+1}/Q_n = [(F_nL_n + 1)L_n^{-1}(L_n + 1)^2 \times (L_n^{-1} + F_n^{-1})]^{N_n/8}. \tag{3c}$$

These equations are not as formidable as they look. First, observe that F_{n+1} and L_{n+1} are uniquely determined by F_n and L_n. Furthermore, their determination does not involve the function Q_n. We may therefore concentrate on Eqs. (3a) and (3b). We first set

$$\alpha_n = (F_n/L_n)^{1/2} \quad [= \exp(2K_n - B_n), \text{ see Appendix A}], \tag{4}$$

$$\beta_n = (F_nL_n)^{1/2} \quad [= \exp(2K_n + B_n), \text{ see Appendix A}].$$

These substitutions bring (3a) and (3b) to the symmetric forms

$$\alpha_{n+1} = (\alpha_n^2 + 1)(\alpha_n + \beta_n)^{-1},$$
$$\beta_{n+1} = (\beta_n^2 + 1)(\alpha_n + \beta_n)^{-1}. \tag{5}$$

There is no sign ambiguity in extracting square roots, as all exponentials are positive. Now *observe that*

$$\beta_{n+1} - \alpha_{n+1} = \beta_n - \alpha_n = \lambda \quad \text{(say)}, \tag{6}$$

which is independent of n; the symmetry of (5) means that there remains only *one* recursion relation to solve. Now define

$$\zeta^2 = 1 + \lambda^2/4, \tag{7}$$

$$\alpha_n = \zeta\delta_n - \lambda/2. \tag{8a}$$

Then

$$\beta_n = \zeta\delta_n + \lambda/2, \tag{8b}$$

and from (5) one obtains the recursion relation for δ_n:

$$\delta_{n+1} = (\delta_n^2 + 1)/2\delta_n. \tag{9}$$

This equation may be solved by an "N/D procedure." Set

$$\delta_n = P_n/D_n, \tag{10}$$

where P_n and D_n have no common divisors. Then (9) becomes

$$\frac{P_{n+1}}{D_{n+1}} = \frac{P_n^2/D_n^2 + 1}{2P_n/D_n} = \frac{P_n^2 + D_n^2}{2P_nD_n},$$

whence, by the hypotheses on P_n and D_n,

$$P_{n+1} = P_n^2 + D_n^2, \quad D_{n+1} = 2P_nD_n \tag{11}$$

These equations are finally decoupled by the substitutions

 0022-2488/79/081759-02$01.00

$$G_n = P_n + D_n, \quad H_n = P_n - D_n. \tag{12}$$

With these substitutions, Eqs. (11) become [3]

$$G_{n+1} = G_n^2, \quad H_{n+1} = H_n^2, \tag{13}$$

which have the unique solutions

$$G_n = G_0^{2^n}, \quad H_n = H_0^{2^n}. \tag{14}$$

3. CONCLUSION

It is now a matter of substituting backwards to obtain explicit expressions for F_n and L_n, and then for the right-hand side of Eq. (3c). Some computation is required to assemble the known final result on the one-dimensional Ising model. [4] This is briefly indicated in Appendix B. More complicated systems of coupled nonlinear recursion relations are currently under study.

APPENDIX A

Equations (3) are obtained for a linear chain of N spins, $N = 2^{M+1}$. The partition function is

$$Z = \sum_{\substack{S_i = \pm 1 \\ i = 1,\cdots N}} \exp\left(A_0 + K_0 \sum_{i=1}^{N} S_i S_{i+1} + B_0 \sum_{i=1}^{N} S_i\right), \tag{A1}$$

with $S_{N+1} = S_1$. In the above $A_0 = 0$ and has been written in for symmetry. We write this as

$$Z = \sum U(A_0, K_0, B_0; S_i, N), \tag{A2}$$

where the spins S_i have also been indicated explicitly. The summation extends over all spin variables S_i, and for typographic convenience we have suppressed the range of summation. The RG transformation consists of summing over alternate spins. After each such transformation, the surviving spins are renumbered serially from 1 to $N/2^m = N_m$. Here m denotes the number of iterations, or the number of times the RG transformation has been performed. After n iterations we obtain

$$Z = \sum U(A_n, K_n, B_n; S_i, N_n). \tag{A3}$$

To write the recursion relations for A_n, K_n and B_n in "algebraic" form we introduce the notations

$$F_n = \exp 4K_n, \quad L_n = \exp 2B_n, \quad Q_n = \exp A_n. \tag{A4}$$

Equations (3) follow immediately. The latter are the same as Eqs. (3.10)–(3.12) of Nelson and Fisher, [1] who solved them numerically for the *infinite* case. Their notations are related to our by

$$x^{-1} = F_n, \quad y^{-1} = B_n, \quad w^{N_n/4} = Q_n, \tag{A5}$$

and they denote the quantities indexed by $n+1$ by primes. Observe that all quantities are positive.

APPENDIX B

Suppose that we start with $N = 2^{M+1}$ spins. Then the RG tranformation can be applied exactly M times. After M iterations, we shall obtain

$$Z = \sum_{S_1, S_2 = \pm 1} \exp[A_M + K_m(S_1 S_2 + S_2 S_1) + B_M(S_1 + S_2)].$$

Summation over the two surviving spins yield

$$Z = 2\zeta \delta_{M+1} \exp(A_{M+1} - K_{M+1}). \tag{B1}$$

The quantities A_{M+1}, B_{M+1}, and K_{M+1} are determined by the same recursion relations (3), in the notations (A4), although the RG transformation cannot be iterated the $M+1$th time.

In order to obtain the partition function in the desired form, we use (10), (12), and (14) to obtain

$$\delta_n = \frac{(\delta_0 + 1)^{N_n} + (\delta_0 - 1)^{N_n}}{(\delta_0 + 1)^{N_n} - (\delta_0 - 1)^{N_n}}, \tag{B2}$$

where we have used $D_0 = 1$, and δ_0 is given by

$$\delta_0 = \exp(2K)\cosh(B)\zeta^{-1}. \tag{B3}$$

Using (A4), (3), and $\lambda = \beta_0 - \alpha_0 = \exp(2K + B) - \exp(2K - B)$ we obtain for the term $\exp A_{M+1}$ in (B1) the expression

$$\exp A_{M+1} = (2\zeta)^{1+2+\cdots+N/2} \frac{\alpha_{M+1}^{1/4} \beta_{M+1}^{1/4}}{(\alpha_0 \beta_0)^{N/4}} \times \delta_M \delta_{M-1}^2 \cdots \delta_0^{N/2}. \tag{B4}$$

Multiplying and dividing by δ_{M+1} and using (10) and (11) we have

$$\delta_M \delta_{M-1}^2 \cdots \delta_0^{N/2} = \frac{P_{M+1}}{D_{M+1}} \frac{1}{2^{1+2+\cdots+N/2} D_0^N}. \tag{B5}$$

Inserting (B4) and (B5) into (B1) we obtain

$$Z = 2\zeta^N P_{M+1}/\exp(NK) = \lambda_+^N + \lambda_-^N, \tag{B6}$$

where

$$\lambda_\pm = e^K[\cosh B \pm (\cosh^2 B - 2\exp(-2K)\sinh 2K)^{1/2}],$$

which is the known result. [4]

[1] D.R. Nelson and M.E. Fisher, Ann. Phys. **91**, 226 (1975). This article gives a relevant review of the RG formalism.

[2] For an early general review of Ising models, see G.F. Newell and E.W. Montroll, Rev. Mod. Phys. **25**, 353 (1953).

[3] It may be of some interest to note that the same questions are finally obtained in the much simpler case of no external field; A. Rabinovitch and R.N. Sen, Ann. Israel Phys. Soc. **2**, (2), 428 (1978); Proceedings of STATPHYS 13 (Adam Hilger and Isreal Physical society).

[4] K. Huang, *Statistical Mechanics* (Wiley, New York, 1967).

JMP0800I JMAPA,00AUG9790806

Moment-method in a hard-sphere problem on a lattice, and physical applications. I. Uniform chain

J.-P. Gallinar

Departamento de Física, Universidad Simón Bolívar, Apartado 80659, Caracas 108, Venezuela
(Received 22 February 1978; revised manuscript received 11 October 1978)

An exact representation is obtained for the propagator of two "hard spheres" on a one-dimensional (uniform) lattice where only nearest-neighbor hops are considered. The problem is solved within a random-walk formulation, with a general expression for the nth moment of the spectral density being given. These formal results are then applied to the discussion of several physical applications in a unified manner. In particular, we show that within the framework of the Hubbard Hamiltonian one may obtain, from the appropriate generalized spectral density, the optical absorption spectrum of a strongly correlated one-dimensional band of electrons. This theory, as applied, for example, to magnetic insulators or to the TCNQ (tetra-cyanoquinodimethane) salts, *predicts* a logarithmic divergence of the optical spectrum, for which a novel interpretation is given here in terms of an equivalent "surface" problem, shown to be isomorphic to the one studied. This is the problem of a two-dimensional rectangular crystal which has been cleaved along a main-diagonal line "surface." From this, we also indicate the effects of chain dimerization (or Peierls transition) in a simple case of the spectral density. A simple counter example is also given which shows the nonuniqueness of the moment-method reconstruction for functions of unbounded variation. The partition function is also obtained analytically, and from this follows the energy of the two holes one-dimensional spinless fermion band. Finally, a physical argument is presented which supports the spinless fermion prediction of infinite mobility (or d.c. conductivity) for the hard spheres.

1. INTRODUCTION

Recently,[1] there has been renewed interest in the properties of a collection of particles or hard spheres that move on a one-dimensional lattice by nearest-neighbor hopping. Such systems have been usually studied either by mean-field theories or by computer simulations and phenomenological equations. The use of moment[2,3] expansions seems to be rather new[1] in the study of such problems. Hence, it seems worthwhile as a first step towards an *exact* treatment of such systems, by infinite moment expansions, to present a solution for a simple two hard-sphere situation on a uniform chain. An exact, closed form representation of the propagator on the uniform chain is obtained in this case, and shown to have useful physical applications. The treatment of "excitonic" effects (longer range interaction) between the hard spheres, and of a dimerized chain, will be the subject of another work, although a very simple case of the latter one is also given here.

The plan of this paper is then as follows: In Sec. 2 the moments of the spectral density of the propagator are calculated by a random-walk technique applied to the *two* hard spheres; we then introduce in Sec. 3 an integral representation that allows here an exact summation of the moment series. Finally, in Sec. 4 several physically interesting applications of our results are discussed in a unified manner. Physically, in these presented applications, the hard spheres can be either electrons, holes, or spinless sites formed by two paired (up spin and down spin) electrons. To conclude, in Appendix C a continued fraction representation is given for a very simple case of the dimerized chain.

2. MOMENTS OF THE SPECTRAL DENSITY

We consider the propagator or resolvent

$$G(z) = (z - \mathscr{H})^{-1} \tag{1}$$

of the Hamiltonian $\mathscr{H}$. This last one will be taken to describe hopping only by nearest neighbors of the two hard spheres. Because of this there is no spatial reordering of the spheres in one dimension. The asymptotic expansion of (1) is then written as[2]

$$G(z) = \frac{\mu_0}{z} + \frac{\mu_1}{z^2} + \frac{\mu_2}{z^3} + \cdots \tag{2}$$

where $\mu_n = \mathscr{H}^n$. To calculate the μ_n's, or moments, we shall choose a basis set $\Psi(p,q)$ in the position representation, such that $\Psi(p,q) \equiv |p,q\rangle$ represents a state vector with a sphere on the lattice point p and with the other on site q; we set $p < q$. The state $\Psi(p,p)$ is projected out because of the hard-core interaction between the spheres. The general matrix element of (1) is then

[a]Supported in part by the Venezuelan Government Agency Consejo Nacional de Investigaciones Científicas y Tecnológicas CONICIT-Project No. 51-26-S1-939.

 0022-2488/79/081761-08$01.00

$$\langle p',q'|G(z)|p,q\rangle = \sum_{n=0}^{\infty} \frac{\langle p',q'|\mathscr{H}^n|p,q\rangle}{z^{n+1}}. \quad (3)$$

To evaluate $\langle p',q'|\mathscr{H}^n|p,q\rangle$ we proceed as follows. One has as definition

$$\langle p',q'|\mathscr{H}|p,q\rangle = T(\delta_{p',p+1}\delta_{q',q} + \delta_{p',p}\delta_{q',q+1} + \delta_{p',p-1}\delta_{q',q} + \delta_{p',p}\delta_{q',q-1}), \quad (4)$$

where T is the magnitude of the nonvanishing matrix element. For $n>1$, the matrix element of $\mathscr{H}^n$ is composed of all processes in which a total of n hops is distributed in all possible manners among the two hard spheres. Thus, one has

$$\mathscr{H}^n|p,q\rangle = T^n \sum_{q_1,q_2} P_n^{(p,q)}(q_1,q_2)|q_1,q_2\rangle, \quad (5)$$

where $P_n^{(p,q)}(q_1,q_2)$ is the number of random walks of n steps that start from $|p,q\rangle$ and end up at $|q_1,q_2\rangle$, with $q_1<q_2$. the quantity $P_n^{(p,q)}(q_1,q_2)$ can then be computed by considering first the sphere at q to be fixed, and allow it at p to move to a position $q+q_1-q_2<q$. For this situation the distance between the spheres is then of q_2-q_1 steps. The number of these particular random walks of n steps is given by[4]

$$\mathscr{P}_n^{(p,q)}(q_1,q_2) = P_n[q_2-q_1-(q-p)] - P_n[q_2-q_1+(q-p)], \quad (6)$$

where

$$P_n[x] \equiv \binom{n}{\frac{1}{2}(n+x)} = \frac{n!}{[\frac{1}{2}(n+x)]![\frac{1}{2}(n-x)]!}. \quad (7)$$

This is the expression[4] for one-dimensional walks against the absorbing barrier of the sphere held fixed at q. By now letting this sphere move to site q_2 and that at $q+q_1-q_2$ to site q_1, one obtains the final state $|q_1,q_2\rangle$, for which the distance between the spheres is also of q_2-q_1 steps, as in the state $|q+q_1-q_2,q\rangle$. The quantity $P_n^{(p,q)}(q_1,q_2)$ can then be obtained from $\mathscr{P}_n^{(p,q)}(q_1,q_2)$ in the following manner: One substitutes in all possible ways k of the n_1 steps taken to the right by the sphere that moved from site p to site $q+q_1-q_2$ by steps taken to the left by the sphere held fixed at q, and also substitutes $k+q_2-q$ of the n_2 steps taken to the left by the sphere that moved from site p to site $q+q_1-q_2$ by steps taken to the right by the sphere held fixed at q. The sphere originally fixed at q obviously ends up at site q_2, while the sphere at $q+q_1-q_2$ ends up at q_1, since the process of substitution does not alter the relative distance between the spheres. We will then have

$$P_n^{(p,q)}(q_1,q_2) = \mathscr{P}_n^{(p,q)}(q_1,q_2)\sum_k \binom{n_1}{k}\binom{n_2}{k+q_2-q}, \quad (8)$$

where

$$n_1+n_2=n \quad \text{and} \quad n_1-n_2=q+q_1-q_2-p, \quad (9)$$

the sum on k being over all possible values. The sum over k can be easily evaluated since one has

$$\sum_k \binom{n_1}{k}\binom{n_2}{k+q_2-q} = \frac{1}{2\pi i}\oint dz\, z^{q_2-q-1}\sum_{k,k'} z^{k-k'}\binom{n_1}{k}\binom{n_2}{k'}$$
$$= \frac{1}{2\pi i}\oint \frac{dz}{z} z^{q_2-q}(1+z)^{n_1}(1+z^{-1})^{n_2}$$
$$= \frac{1}{2\pi i}\oint \frac{dz}{z} z^{q_2-q-n_2}(1+z)^{n_1+n_2}$$
$$= \sum_k \frac{1}{2\pi i}\oint \frac{dz}{z} z^{q_2-q-n_2+k}\binom{n}{k}$$
$$= \binom{n}{n_2+q-q_2}, \quad (10)$$

where the counterclockwise contour with $|z|<1$ need only enclose the origin in the z plane. Solving (9) for n_2, one finds that

$$n_2+q-q_2 = \tfrac{1}{2}(n-q_1-q_2+q+p) \quad (11)$$

so that finally

$$\sum_k \binom{n_1}{k}\binom{n_2}{k+q_2-q} = P_n[q_1+q_2-(q+p)] \quad (12)$$

and

$$P_n^{(p,q)}(q_1,q_2) = \mathscr{P}_n^{(p,q)}(q_1,q_2)P_n[q_1+q_2-(q+p)]$$
$$= (P_n[q_2-q_1-(q-p)] - P_n[q_2-q_1+(q-p)])P_n[q_1+q_2-(q+p)], \quad (13)$$

which then gives the generalized nth moment of the spectral density.[2]

One then has

$$\langle p',q'|\mu_n|p,q\rangle \equiv \langle p',q'|\mathscr{H}^n|p,q\rangle = T^n P_n^{(p,q)}(p',q') \quad (14)$$

and

$$\langle p',q'|G(z)|p,q\rangle = \sum_{n=0}^{\infty} \frac{T^n}{z^{n+1}} P_n^{(p,q)}(p',q'), \quad (15)$$

the series in (15) converging for $|z|$ sufficiently large. This is evaluated in the next section in terms of an integral representation and the region of convergence is established.

3. INTEGRAL REPRESENTATION OF THE PROPAGATOR AND SPECTRAL DENSITY

Defining $T/z = \mathfrak{z}$, we obtain

$$\langle p',q'|G(z)|p,q\rangle = \frac{1}{z}\sum_{n=0}^{\infty} \mathfrak{z}^n P_n^{(p,q)}(p',q'), \quad (16)$$

where we evaluate the series in n by introducing an integral representation for $P_n^{(p,q)}(p',q')$, such that[4]

$$P_n^{(p,q)}(p',q') = (P_n[q'-p'-(q-p)] - P_n[q'-p'+(q-p)]) \times \frac{1}{2\pi i}\oint \frac{dz'}{z'^{n+1}}\left(\frac{1-(1-4z'^2)^{1/2}}{2z'}\right)^{|p'+q'-(q+p)|} \times (1-4z'^2)^{-1/2}. \quad (17)$$

The contour of integration in (17), as shown in Fig. 1, is a counterclockwise contour C_1 with $|z'|<\frac{1}{2}$ which encloses the origin $z'=0$, and in (17) we have written for the second factor of (13) an integral representation.[4]

One now deforms the contour C_1 to that of C_2 around the cuts on the real axis, which extend from $|z'|=\frac{1}{2}$ to $|z'|=\infty$. Since the integrand obviously vanishes faster than z'^{-1}, the integral on the circle at infinity is zero. By flatten-

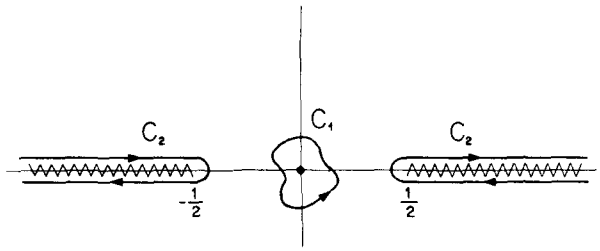

FIG. 1. Contours used in the evaluation of Eq. (17).

ing out the contour C_2 one is finally left with an integral of the discontinuity of the integrand across the cut, which is simply given by its imaginary part. We will then have that

$$\mathrm{Im}\left(\frac{1-(1-4z'^2)^{1/2}}{2z'}\right)^{|p'+q'-(q+p)|}(1-4z'^2)^{-1/2}$$
$$=(4z'^2-1)^{-1/2}\,\mathrm{Re}\left(\frac{1+i(4z'^2-1)^{1/2}}{2z'}\right)^{|p'+q'-(q+p)|} \tag{18}$$

But since $1+i(4z'^2-1)^{1/2}=2z'\exp(i\theta)$ where $\tan\theta=(4z'^2-1)^{1/2}$ and $\cos\theta=1/2z'$, one has that

$$\mathrm{Re}\left(\frac{1+i(4z'^2-1)^{1/2}}{2z'}\right)^{|p'+q'-(q+p)|}$$
$$=\cos|p'+q'-(q+p)|\theta \tag{19}$$

so that

$$P_n^{(p,q)}(p',q')=\frac{1}{2\pi i}2\int_{1/2}^{\infty}\frac{dz'}{z'^{n+1}}(P_n[q'-p'-(q-p)]$$
$$-P_n[q'-p'+(q-p)])(2i)$$
$$\times\frac{\cos|p'+q'-(q+p)|\theta}{(4z'^2-1)^{1/2}} \tag{20}$$

for n of the same parity as $p'+q'-(q+p)$, and zero otherwise. The integral representation of the propagator is then given by

$$\langle p',q'|G(z)|p,q\rangle$$
$$=\frac{2}{\pi z}\int_{1/2}^{\infty}\frac{dz'}{(4z'^2-1)^{1/2}}\cos|p'+q'-(q+p)|$$
$$\times\sum_{n=0}^{\infty}\frac{\not{z}^n}{z'^{n+1}}(P_n[q'-p'-(q-p)]$$
$$-P_n[q'-p'+(q-p)]), \tag{21}$$

where the orders of integration and summation have been interchanged. The sum on n can be easily done since it can be expressed in terms of the generating function[4] for one-dimensional random walks. In fact, for $|\not{z}/z'|<\frac{1}{2}$

$$\sum_{n=0}^{\infty}\frac{\not{z}^n}{z'^{n+1}}P_n[q'-p'-(q-p)]$$
$$=\left(\frac{1-[1-4(\not{z}/z)^2]^{1/2}}{2(\not{z}/z')}\right)^{|q'-p'-(q-p)|}$$
$$\times[1-4(\not{z}/z')^2]^{-1/2} \tag{22}$$

so that finally

$$\langle p',q'|G(z)|p,q\rangle$$
$$=\frac{2}{\pi z}\int_{1/2}^{\infty}\frac{dz'}{(4z'^2-1)^{1/2}}\frac{\cos|p'+q'-(q+p)|\theta}{z'}$$
$$\times[1-4(\not{z}/z')^2]^{-1/2}$$
$$\times\left[\left(\frac{1-[1-4(\not{z}/z')^2]^{1/2}}{2(\not{z}/z')}\right)^{|q'-p'-(q-p)|}\right.$$
$$\left.-\left(\frac{1-[1-4(\not{z}/z')^2]^{1/2}}{2(\not{z}/z')}\right)^{|q'-p'+(q-p)|}\right] \tag{23}$$

and there are no vanishing matrix elements, since the relative parity of the numbers $p'+q'-(q+p)$, $q'-p'-(q-p)$, and $q'-p'+(q-p)$ is the same. The region of convergence of the series representation in (15) is easily given from the integral representation (23). In fact, this last one represents $G(z)$ everywhere in the complex z plane, and is the unique analytic continuation of (15) inside the circle $|z|=4T$, having an infinite number of identical moments as (15) in the region $|z|>4T$, where (15) converges. The function in (23) has a cut on the real axis inside the circle $|z|=4T$, where the spectral density will not vanish. In fact, the process of analytic continuation represented in going from (15) to (23) is fundamentally the same as that of the continued fraction[2] obtained as the $(n\to\infty)$ limit of the rational $[n,n-1](z)$ Padé[5,6] approximants of (15). In the latter case the poles and zeros of the Padé approximants become dense, as $n\to\infty$, in the interval $(-4T,4T)$ where the cut of the function (23) appears.

The diagonal element when the spheres start some distance ν apart, i.e., when $p'=p$ and $q'=q$ with $q=p+\nu$, is associated[2] with a usual diagonal spectral density. Denoting this matrix element by $G^{(\nu)}(z)$, we obtain

$$G^{(\nu)}(z)=\frac{2}{\pi z}\int_{1/2}^{\infty}\frac{dz'}{(4z'^2-1)^{1/2}}(z'^2-4\not{z}^2)^{-1/2}$$
$$\times\left[1-\left(\frac{1-[1-4(\not{z}/z')^2]^{1/2}}{2(\not{z}/z')}\right)^{2\nu}\right]. \tag{24}$$

The corresponding spectral density is given by[2]

$$\mathscr{G}^{(\nu)}(\epsilon)=\frac{1}{2\pi i}[G^{(\nu)}(\epsilon-i\delta)-G^{(\nu)}(\epsilon+i\delta)]$$
$$=\frac{1}{\pi}\mathrm{Im}G^{(\nu)}(\epsilon-i\delta) \tag{25}$$

which in turn gives back the associated diagonal moments, since

$$\mu_n^{(\nu)}\equiv\langle p,p+\nu|\mathscr{H}^n|p,p+\nu\rangle=\int_{-\infty}^{+\infty}d\epsilon\,\epsilon^n\,\mathscr{G}^{(\nu)}(\epsilon) \tag{26}$$

clearly holds.

From (24) and (25) one finds easily the discontinuity of $G^{(\nu)}(z)$ across the cut on the real axis; this is, of course, possible since in the integral representation (24) *all*[2] the moments of $G^{(\nu)}(z)$ are implicitly given. One then finds, for $|z|\leqslant 4T$,

$$\mathscr{G}^{(\nu)}(z)=\frac{2}{\pi^2}\int_{1/2}^{2T/|z|}\frac{dz'}{(4z'^2-1)^{1/2}}\mathrm{Im}[(z'^2z^2-4T^2)^{-1/2}]$$
$$\times\left[1-\left(\frac{z'z-(z'^2z^2-4T^2)^{1/2}}{2T}\right)^{2\nu}\right] \tag{27}$$
$$=\frac{2}{\pi^2}\int_{1/2}^{2T/|z|}\frac{dz'}{(4z'^2-1)^{1/2}}(4T^2-z'^2z^2)^{-1/2}$$

$$\times \mathrm{Re}\left[1-\left(\frac{z'z+i(4T^2-z'^2z^2)^{1/2}}{2T}\right)^{2\nu}\right]. \tag{28}$$

But

$$\mathrm{Re}\left(\frac{z'z+i(4T^2-z'^2z^2)^{1/2}}{2T}\right)^{2\nu}=\cos 2\nu\alpha, \tag{29}$$

where $\cos\alpha = z'z/2T$, and so

$$\mathcal{G}^{(\nu)}(z)=\frac{4}{\pi^2}\int_{1/2}^{2T/|z|}\frac{dz'}{(4z'^2-1)^{1/2}}\frac{\sin^2\nu\alpha}{(4T^2-z'^2z^2)^{1/2}} \tag{30}$$

for $|z|\leqslant 4T$ and zero otherwise. This can be written in a more convenient form by changing variables to $k'\equiv|z|/4T\leqslant 1$ and $x=2k'z'\leqslant 1$, so that

$$\mathcal{G}^{(\nu)}(z)=\frac{1}{\pi^2 T}\int_{k'}^{1}\frac{dx}{(x^2-k'^2)^{1/2}}\frac{\sin^2\nu\alpha}{(1-x^2)^{1/2}}, \tag{31}$$

where $\cos\alpha = x$. Since $\sin^2\nu\alpha$ can always be expressed as a rational function of x, the integral in (31) can always be expressed[7] in terms of the complete elliptical integrals $K(k)$, $E(k)$ of the first and second kind, respectively.

For $\nu = 1$,

$$\begin{aligned}\mathcal{G}^{(1)}(z)&=\frac{1}{\pi^2 T}\int_{k'}^{1}\frac{(1-x^2)^{1/2}}{(x^2-k'^2)^{1/2}}\\&=\frac{1}{\pi^2 T}[K(k)-E(k)],\end{aligned} \tag{32}$$

where $k=(1-k'^2)^{1/2}\leqslant 1$ is the modulus of the integrals. For $\nu = 2$,

$$\begin{aligned}\mathcal{G}^{(2)}(z)&=\frac{4}{\pi^2 T}\int_{k'}^{1}dx\,\frac{x^2(1-x^2)^{1/2}}{(x^2-k'^2)^{1/2}}\\&=\frac{4}{3\pi^2 T}\{(1-k'^2)E(k)+k'^2[K(k)-E(k)]\}.\end{aligned} \tag{33}$$

In general for ν odd $\mathcal{G}^{(\nu)}(z)$ diverges logarithmically for $z\to 0$, while for ν even $\mathcal{G}^{(\nu)}(z=0)$ remains finite but increases with ν, so that in general

$$\lim_{\nu\to\infty}\mathcal{G}^{(\nu)}(z=0)=\infty. \tag{34}$$

Furthermore, when $\nu\to\infty$, $\mathcal{G}^{(\nu)}(z)$ converges to

$$\begin{aligned}\mathcal{G}^{(\infty)}(z)&=\frac{1}{2\pi^2 T}\int_{k'}^{1}\frac{dx}{(x^2-k'^2)^{1/2}}(1-x^2)^{-1/2}\\&=K(k)/2\pi^2 T\end{aligned} \tag{35}$$

everywhere except at $k' = 1$.

Let us turn now to a transformed generating function[4] $G_\alpha(z)$ for the two-particle random walk. We notice that (15) can be rewritten as

$$\langle\bar p',\bar q'|\,G(z)\,|\bar p,\bar q\rangle=\sum_{n=0}^{\infty}\frac{T^n}{z^{n+1}}\,P_n^{(\bar p,\bar q)}(\bar p',\bar q'), \tag{36}$$

where $\bar p = q-p$, $\bar q = q+p$, $\bar p' = q'-p'$, $\bar q' = q'+p'$, and

$$P_n^{(\bar p,\bar q)}(\bar p',\bar q')=(P_n[\bar p'-\bar p]-P_n[\bar p'+\bar p])P_n[\bar q'-\bar q]. \tag{37}$$

We then introduce[4]

$$\langle\bar p'|\,G_\alpha(z)\,|\bar p\rangle\equiv\sum_{n=0}^{\infty}\frac{T^n}{z^{n+1}}(P_n[\bar p'-\bar p]-P_n[\bar p'+\bar p])$$

$$\times\sum_{\bar q'}\alpha^{|\bar q'-\bar q|}P_n[\bar q'-\bar q], \tag{38}$$

where $|\alpha|\leqslant 1$. The sum over $\bar q'$ can be evaluated[8,9] in terms of the integral representation introduced in (17). One finds[9] for $\bar p'-\bar p$ even,

$$\begin{aligned}\sum_{\bar q'}\alpha^{|\bar q'-\bar q|}P_n[\bar q'-\bar q]&=\frac{2}{\pi}\int_{1/2}^{\infty}\frac{dx}{x^{n+1}}(4x^2-1)^{-1/2}\\&\times\left(\frac{x^2(1-\alpha^4)}{x^2(1+\alpha^2)^2-\alpha^2}\right),\end{aligned} \tag{39}$$

while for $\bar p'-\bar p$ or $\bar q'-\bar q$ odd, one finds (see Appendix A)

$$\begin{aligned}\sum_{\bar q'}\alpha^{|\bar q'-\bar q|}P_n[\bar q'-\bar q]&=\frac{2\alpha}{\pi}\int_{1/2}^{\infty}\frac{dx}{x^{n+1}}\frac{x}{(4x^2-1)^{1/2}}\\&\frac{(1-\alpha^2)}{x^2(1+\alpha^2)^2-\alpha^2}.\end{aligned} \tag{40}$$

Then for $\bar p'-\bar p$ even, (38) becomes

$$\begin{aligned}&\langle\bar p'|\,G_\alpha(z)\,|\bar p\rangle\\&=\frac{2}{\pi z}\int_{1/2}^{\infty}\frac{dx}{x(4x^2-1)^{1/2}}\left(\frac{x^2(1-\alpha^4)}{x^2(1+\alpha^2)^2-\alpha^2}\right)\\&\times\sum_{n=0}^{\infty}\frac{T^n}{(xz)^n}(P_n[\bar p'-\bar p]-P_n[\bar p'+\bar p]),\end{aligned} \tag{41}$$

the sum on n being evaluated according to Eq. (22). We note that (41) can also be obtained directly from (23), since

$$\sum_{q(\mathrm{even})}\alpha^{|q|}\cos|q|\theta=\frac{1-\alpha^4}{1+\alpha^4-2\alpha^2\cos 2\theta} \tag{42}$$

with $\cos\theta = (2x)^{-1}$.

One may consider the spectral density associated with (41) when $\alpha = 1$, here (42) behaves as $\delta(\theta)$, and from (38) one obtains

$$\langle\bar p'|\,G_{\alpha=1}(z)\,|\bar p\rangle=\sum_{n=0}^{\infty}\frac{(2T)^n}{z^{n+1}}(P_n[\bar p'-\bar p]-P_n[\bar p'+\bar p]). \tag{43}$$

Then for $\bar p'=\bar p=\nu$, one finds

$$\mathrm{Im}\langle\bar p|\,G_{\alpha=1}(z)\,|\bar p\rangle=2\sin^2\nu\theta/[(4T)^2-z^2]^{1/2} \tag{44}$$

with $\cos\theta = z/4T$. We will turn now to several physical applications of the formal results obtained in Secs. 2 and 3.

4. PHYSICAL APPLICATIONS AND COMMENTS

In the preceding Secs. 2 and 3 we have established some formal results pertaining to the propagator and some of its generalized spectral densities, of a Hamiltonian $\mathcal{H}$ that describes a system of two hard spheres that are allowed to hop on a chain only through nearest neighbors. The chain is described by a single matrix element T between nearest neighbors. Some direct physical applications of these results will be made in this section; as we will show, some of the spectral densities obtained can be directly related to *experimentally* measurable quantities. Let us first note how the system treated is mathematically isomorphic to a different physical problem. This is the problem[10] (see Fig. 2) of a two-dimensional rectangular crystal which has been cleaved along a main diagonal line "surface."

Conceptually, this can be visualized either as a real[10] crystal in which three-dimensional aspects may be neglected

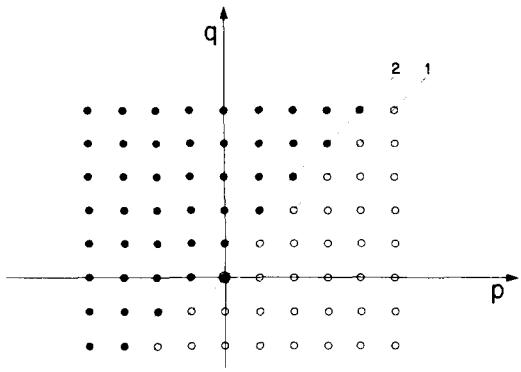

FIG. 2. Two-dimensional lattice phase space. Absorbing barriers on line 1. Reflecting barriers on line 2. Origin and hollow dots cannot be reached by the representative point.

(such as might be the case for states[10] localized near the two-dimensional surface of the cleaved crystal) or as a two-dimensional lattice phase space where the representative point of the two hard spheres moves. According to the matrix elements in (4) the representative point is allowed to execute a random-walk motion only between nearest-neighbor lattice sites in the plane lattice, with a matrix element for hopping also given by T. The hard-core condition on the spheres is constructed by cleaving the two-dimensional crystal and introducing a line surface of absorbing[4] barriers at the diagonal points $q = p$. This is represented by line 1 in Fig. 2. As pointed out by Di Marzio,[11] such a line of absorbing barriers is equivalent physically to an adjoining parallel line (line 2 in Fig. 2) of perfectly reflecting[4] barriers. Since in one dimension the hard-spheres cannot be reordered by nearest-neighbor hopping, the points in phase space (Fig. 2) denoted by empty circles cannot be reached by the representative point. Also, we note that the transformation from the q and p axes to the $\bar{q}$ and $\bar{p}$ variables [Eq. (36)] presented in Sec. 3 amounts then to a $\pi/4$ clockwise rotation about the origin in this phase space. Within this context, the spectral density $\mathcal{G}^{(\nu)}(z)$ in (31) gives then the *local density of states* (see Ref. 10) on the $(\nu + 1)$th crystallographic line parallel to the line surface 1. As $\nu \to \infty$ the local density of states increasingly oscillates in a damped manner, converging to $\mathcal{G}^{(\infty)}(z)$, the uncleaved crystal density of states. Furthermore, we note that the N hard-spheres generalization of this isomorphism (with $N \geqslant 2$) involves a N-dimensional cubic lattice, where the constraints $x_1 < x_2 < x_3 < \cdots < x_N$ on the positions of the spheres, are given by the intersection of appropriate hypersurfaces.

Except for $N = 2$, however, the resulting phase space will always have wedges and corners. For this case, the corresponding exact solution has not been obtained.

Let us consider now the electronic band of spinless fermions in one dimension. Even though electrons have spins, one can show[12,13] that for a strongly correlated unidimensional tight-binding band, such as can be described by a many-body Hamiltonian, like the Hubbard Hamiltonian[14] with $U = \infty$ (U giving the energy of repulsion of two electrons of opposite spins on the *same* lattice site) and with nearest-neighbors hops, the magnetic properties become those of a set of localized spins, completely decoupled from the kinetic degrees of freedom and given by a Curie spin susceptibility. The origin of this somewhat paradoxical behavior lies in the fact that in one dimension, nearest-neighbor hops alone cannot reorder[13] the electronic spins; thus simulating the hard-sphere behavior. The partition function[15] of such a highly interacting system becomes that of a spinless fermion simple tight-binding band with parameter T. From our results a partition function Z can be obtained in a manner similar to the path-counting technique of Beni *et al.*[15] In fact, Z will be given by

$$Z(\beta) = N \sum_{\nu=1}^{N} \int dz\, e^{-\beta z} \mathcal{G}^{(\nu)}(z), \tag{45}$$

where $\mathcal{G}^{(\nu)}(z)$ gives the local density of states at the $(\nu + 1)$th crystallographic line, and $N \to \infty$ is the manner of lattice sites. Then one obtains, interchanging summation with integration

$$Z(\beta) = \frac{N}{\pi^2 T} \int dz\, e^{-\beta z} \int_{k'}^{1} \frac{dx}{(x^2 - k'^2)^{1/2}} \times (1 - x^2)^{-1/2} \sum_{\nu=1}^{N} \sin^2 \nu\alpha, \tag{46}$$

but

$$\sum_{\nu=1}^{N} \sin^2 \nu\alpha = \tfrac{1}{2}(N + \tfrac{1}{2} - \sin(2N + 1)\alpha/2 \sin\alpha). \tag{47}$$

Thus, except for $\alpha = 0$, the last term in (47) is thermodynamically negligible when $N \to \infty$, so that finally

$$Z(\beta) = \frac{2N^2}{\pi^2} \int_{-1}^{+1} dk' \exp(-4\beta T k') K((1 - k'^2)^{1/2}) = N^2 I_0^2(2\beta T), \tag{48}$$

where $I_0(2\beta T)$ is the hyperbolic[16] Bessel function of zero order. The internal energy follows from

$$E = -\frac{\partial \ln Z}{\partial \beta} = -4T \frac{I_1(2\beta T)}{I_0(2\beta T)}, \tag{49}$$

which agrees with the result of Beni *et al.*[15] for a density ρ of spinless fermions of $\rho = 1 - n/N$ (i.e., there are n holes in the half-filled band and ρ = number of electrons/number of sites) with $n = 2$. In fact, one may surmise that for $n \ll N$, the partition function Z should be given by

$$Z(\beta) \cong N^n I_0^n(2\beta T), \tag{50}$$

where terms of the order of $n/N \ll 1$ or higher are neglected. This comes about because the reflections of the representative point of the n holes (or hard spheres) against the surfaces of the phase space, are thermodynamically negligible [as in (47)] when $N \to \infty$ and $n/N \ll 1$. Hence, it follows from this simple argument and the electron-hole symmetry, that for a low $n/N \ll 1$ concentration of electrons (or holes, from the half-filled, $\rho = 1$, band), these will seem[17,18] to behave as spinless fermions, giving rise to an absorption rate proportional to a delta function $\delta(\omega)$, where ω is an[19] external frequency. However, for an arbitrary concentration n/N of electrons, there has been no rigorous proof[17,18] of this assertion, and it is certainly not clear from our argument whether the collisions of the representative point against the surfaces of phase space, coming from higher-order terms in n/N, would still yield a conductivity $\sigma(\omega)$ proportional to $\delta(\omega)$, as

such collisions are thermodynamically certainly nonnegligible.

Another useful application of the spectral densities obtained, lies in the study of the optical absorption spectrum[8,9] of a strongly correlated electronic band of the Hubbard type. Although for the spinless fermion band discussed above one has to consider an assembly of many "hard spheres" (represented either by the electrons or the holes), and the condition $U=\infty$ forbids double occupancies of lattice sites by paired electrons, thereby yielding a null conductivity $\sigma(\omega)$ for an *exactly* half-filled ($\rho=1$) band; one may consider a first order correction to this in terms of the parameter $T/U \ll 1$. Here double occupancies are possible since $U \neq \infty$. However, as it has already been shown,[8,9] one may prove then that the optical absorption spectrum ($\omega \sim U$) of the frequency dependent $\sigma(\omega)$ conductivity, is given essentially in terms of the appropriate spectral density of the two "hard spheres" *alone*. This may not seem exceedingly surprising, if one realizes that the optical spectrum of such a many-body system, arises physically (at low temperatures), from the creation by a light quantum, of an excitonlike excitation, composed of a two-electron paired site and the immediately adjacent created hole. Both the doubly occupied site and the hole propagate independently as "hard spheres" in the lattice, until they are annihilated somewhere by mutual recombination. Although in principle it would be possible to have multiple creations and annihilations of other electron-hole excitoniclike pairs, thereby creating in the lattic innumerable hard spheres, such processes are of higher order[8,9] in $T/U \ll 1$, and hence negligible for $U \to \infty$. In fact, as we have indicated previously[9] the line shape of absorption is given by the spectral density of the transformed generating function in (41) for $\nu=1$ (see Appendix B), i.e.,

$$\mathrm{Im}\langle 1|G_\alpha(z)|1\rangle = \frac{1}{\pi T}\Big((1-k'^2\sin^2\beta)\cos\beta\ K(k) + \frac{\pi}{2}(1-k'^2\sin^2\beta)^{1/2} \times \sin\beta\ \Lambda_0(\beta,k) - \cos\beta\ E(k)\Big), \quad (51)$$

where $\Lambda_0(\beta,k)$ is Heuman's lambda function[7] and

$$\beta = \sin^{-1}(2/(\alpha+\alpha^{-1})). \quad (52)$$

In expression (51) the center of the line shape is at $z=0$, and the α parameter of the transformed generating function is (see Ref. 9) seen to be related to the spin configuration of the rest of the electrons. To the lowest order in T/U these can be considered as fixed with a given spin configuration. Here the spin configuration is important, since for $U \neq \infty$ the electronic spins may be reordered by nearest-neighbor hopping. The spectral density (51) has been previously[9] applied to some recently studied[20] low dimensional organic systems of the TCNQ salts, and one-dimensional magnetic insulators, with strongly correlated electronic bands.

It is important to point out here several novel features[8,9] of this spectrum, that can be understood in terms of the cleaved crystal isomorphism. Both the absorption width, extending to $\pm 4T$ on both sides of $k'=0$, and the logarithmic divergence[21] at the center $k'=0$ of the spectrum, are characteristics[10] of the two-dimensional nature of the phase space. Furthermore, the weak residual logarithmic divergence in the spectrum (51), given for α close to one by

$$\lim_{\alpha\to 1}\lim_{z\to 0}\mathrm{Im}\langle 1|G_\alpha(z)|1\rangle = \frac{1}{\pi T}\lim_{\beta\to\pi/2}[(\cos\beta)\ln(4/k')_{k'\to 0} + \beta\sin\beta - \cos\beta], \quad (53)$$

is interpreted as indicative[10] of the existence of *virtual bound states* associated with the line $\nu=1$, or "surface" of the cleaved crystal. The physical reason why one considers only the plane $\nu=1$ in the optical spectrum in (51), is due to the fact that the quantum of light originally creates the excitation composed by the hole and the doubly occupied site on adjacent sites. Thus, the representative point (Fig. 2) starts somewhere on line 2 and propagates freely in the cleaved crystal, until it must come back to some other point on line 2, where, so to speak, it disappears because of mutual recombination of the hole and the electron. The transformed generating function in (38) and its associated spectral density (51) physically represent nothing but a weighted average (given by $\alpha^{|\bar{q}'-\bar{q}|}$) of such decreasing probabilities of recombination of the electron-hole pair, as its representative point is annihilated further away on line 2, from its starting point on the same line.

We should remark that, although for the conductivity $\sigma(\omega)$ or optical absorption only the plane $\nu=1$ is involved, in other transport properties, like the thermopower, additional (i.e., $\nu=2$) planes will certainly be involved; this being caused by the way the original excitation is created. For this other transport property, next-nearest neighbors should also be taken into account.

The ground state of the many-body system of which (51) gives the optical absorption spectrum, is thought to be[8,9] antiferromagnetic. Because of the translational invariance of this spin arrangement, all mutual recombinations along the points of line 2 (Fig. 2) are then equally certain, and one obtains the ground state spectrum with $\alpha=1$ in (51) [or $\nu=1$ in (44)], i.e.,[22]

$$\mathrm{Im}\langle 1|G_{\alpha=1}(z)|1\rangle = [(4T)^2 - z^2]^{1/2}/8T^2. \quad (54)$$

On the other hand, for $\alpha=0$ in (51), the representative point must come back to its starting point to be annihilated, i.e., one obtains the local density of states on the surface of the cleaved crystal, given by (32). It can be shown[9] that this limit $\alpha \to 0$, insofar as it is applied to the optical spectrum discussed, gives the line shape of absorption for a ferromagnetic arrangement of the spins in the lattice. For this spin arrangement the electron-hole excitation cannot recombine anywhere but at its starting point (i.e., $\alpha=0$) in order to restore the initial ferromagnetic arrangement.

It is well known that quasi-one-dimensional systems, such as some TCNQ salts,[20] for example, are thought to be unstable against Peierls[23] distortions.[24] It is interesting to point out that the phase space isomorphism appropriate here for a Peierls-dimerized linear lattice, with two different, T and T', hopping matrix elements, is analogous to the one presented, except that each lattic point in Fig. 2 should be

connected with its nearest-neighbors through matrix elements given by T and T' alternatively, in both the p and the q directions. Starting from this, it is straightforward to obtain the generalization of (54) with $T \neq T'$, because of the essentially one-dimensional nature of the excursions of the representative point in this case (see Appendix C). The general treatment of this for $|\alpha| \leqslant 1$ [for example, appropriate for a general disordered spin arrangement as in (51)] will be the object of a future work (part II).

Finally, we also remark that we have laid the grounwork for the inclusion of longer-range interaction between the hard spheres. For example, the inclusion of nearest-neighbor interactions within the context of an extended[25,26] Hubbard model, may lead here to the formation of bound electron-hole pairs of an excitonic nature, because of the effective attraction between a doubly occupied site and an adjacent hole in this case. Within the discussed isomorphism, this amounts very simply to the inclusion of a constant single-body potential[10] *on* the surface (line 2) of the cleaved crystal. These other important physical applications will also be discussed in Part II.

APPENDIX A

We show Eq. (40) for $\bar{p}' - \bar{p}$ odd; the proof for $\bar{p}' - \bar{p}$ even[9] being completely analogous. Introducing the integral representation of (17), one finds

$$\begin{aligned}\sum_{\bar{q}} \alpha^{|\bar{q}'-\bar{q}|} P_n[\bar{q}'-\bar{q}] &= \frac{1}{2\pi i}\oint \frac{dz}{z^{n+1}}(1-4z^2)^{-1/2} \\ &\times \sum_{\substack{q=-\infty \\ q=\text{odd}}}^{+\infty} \alpha^{|q|}\left(\frac{1-(1-4z^2)^{1/2}}{2z}\right)^{|q|} \\ &= \frac{1}{2\pi i}\oint \frac{dz}{z^{n+1}}(1-4z^2)^{-1/2}\,2\alpha \\ &\times\left(\frac{1-(1-4z^2)^{1/2}}{2z}\right) \\ &\times\left[1-\alpha^2\left(\frac{1-(1-4z^2)^{1/2}}{2z}\right)^2\right]^{-1}, \end{aligned} \tag{A1}$$

where the contour of integration is that of C_1 in Fig. 1. One now deforms the contour to that of C_2 and proceeds as in the proof of Eq. (20). One must notice that the integrand in (A1) has no additional sigularities except the cuts from $|z| = \frac{1}{2}$ to $|z| = \infty$, since the equation $2z = \alpha(1-(1-4z^2)^{1/2})$ has (for $z \neq 0$) only spurious roots at $z = (1/\alpha + \alpha)^{-1}$ for $\alpha < 1$. When $\alpha = 1$ the poles of the integrand in (A1) that appear at $|z| = \frac{1}{2}$ become merged within the branch points also at $|z| = \frac{1}{2}$. One then finds

$$\begin{aligned}\sum_{\bar{q}} \alpha^{|\bar{q}'-\bar{q}|} P_n[\bar{q}'-\bar{q}] &= \frac{4\alpha}{\pi}\int_{1/2}^{\infty} \frac{dx}{x^{n+1}}\frac{2x}{(4x^2-1)^{1/2}} \\ &\times \mathrm{Re}\left(\frac{1+i(4x^2-1)^{1/2}}{4x^2-\alpha^2(1+i(4x^2-1)^{1/2})^2}\right), \end{aligned} \tag{A2}$$

which leads to (40). Notice also that for $\bar{p}' - \bar{p}$ odd, n must be odd, and that, apart from the factor $(1/z)^{n+1}$ in (A1), the discontinuity on the real axis of the rest of the integrand in (A1) is an even function of z.

APPENDIX B

We show Eq. (51). Starting from (41) one obtains

$$\begin{aligned} I &\equiv \pi T\,\mathrm{Im}\langle 1|G_\alpha(z)|1\rangle \\ &= \int_{k'}^{1} dx \frac{(1-x^2)^{1/2}}{(x^2-k'^2)^{1/2}}\frac{x^2(1-\alpha^4)}{x^2(1+\alpha^2)^2-4\alpha^2k'^2} \end{aligned} \tag{B1}$$

and the substitution $x = \cos\theta$, $\sin\psi = \sin\theta/k$ brings (B1) to the form

$$I = \left(\frac{1-\alpha^2}{1+\alpha^2}\right)\int_0^{\pi/2} d\psi \frac{n^2\sin^2\psi(1-k^2\sin^2\psi)^{1/2}}{1-n^2\sin^2\psi}, \tag{B2}$$

where

$$n^2 = \frac{k^2(1+\alpha^2)^2}{(1-\alpha^2)^2+4\alpha^2k^2}, \quad \text{with } 1 \geqslant n^2 \geqslant k^2, \tag{B3}$$

is the parameter of the elliptic integral of the third kind[7] in one of its circular cases because of inequality (B3). Then

$$\begin{aligned} I = \left(\frac{1-\alpha^2}{1+\alpha^2}\right)&[\Pi(n^2,k)(1-k^2/n^2)+(k^2/n^2)K(k) \\ &-E(k)], \end{aligned} \tag{B4}$$

where $\Pi(n^2,k)$ is the complete integral of the third kind. By using the addition formulas[7] for $\Pi(n^2,k)$, one finds

$$\begin{aligned} I = &\frac{(1-\alpha^2)^2+4\alpha^2k^2}{1-\alpha^4}[K(k)-\Pi(-(1-\alpha^2)^2/4\alpha^2,k)] \\ &+\frac{\pi\alpha}{(1+\alpha^2)^2}[(1-\alpha^2)^2+4\alpha^2k^2]^{1/2} \\ &-\left(\frac{1-\alpha^2}{1+\alpha^2}\right)E(k) \end{aligned} \tag{B5}$$

and the evaluation of $\Pi(\alpha_1^2,k)$ with Heuman's lambda function $\Lambda_0(\beta,k)$, through[7]

$$\Pi(\alpha_1^2,k) = \frac{K(k)}{1-\alpha_1^2} + \frac{\pi\alpha_1^2[\Lambda_0(\beta,k)-1]}{2[\alpha_1^2(1-\alpha_1^2)(\alpha_1^2-k^2)]^{1/2}}, \tag{B6}$$

where $\alpha_1^2 \equiv -(1-\alpha^2)^2/4\alpha^2$, and

$$\beta \equiv \sin^{-1}\frac{1}{(1-\alpha_1^2)^{1/2}} = \sin^{-1}\frac{2}{(\alpha+\alpha^{-1})} \tag{B7}$$

finally leads to (51).

APPENDIX C

We generalize (54) to obtain the corresponding spectrum for a Peierls-dimerized linear lattice. Because of the one-dimensional nature of the excursions of the representative point, these can be enumerated by using the continued fraction representation (or infinite order Padé approximant) method employed by Brinkman and Rice.[27] Denoting by T and T' the consecutive matrix elements for hopping on the linear lattice, we then obtain

$\mathrm{Im}\langle 1|G_{\alpha=1}(z)|1\rangle_D$

$$\mathrm{Im}\langle 1|G_{\alpha=1}(z)|1\rangle_D = \mathrm{Im}\cfrac{1}{\left[z\ \ 1-\cfrac{(2T)^2}{\left[z^2\ \ 1-\cfrac{(2T')^2}{z^2\left[1-\cfrac{(2T)^2}{1-\vdots}\right.}\right.}\right.} \tag{C1}$$

This continued fraction can be easily summed in terms of the "self-energies" $\Sigma(z)$ and $\Sigma'(z)$, where

$$\Sigma(z)=\frac{(2T)^2}{z^2[1-\Sigma'(z)]} \quad \text{and} \quad \Sigma'(z)=\frac{(2T')^2}{z^2[1-\Sigma(z)]}. \tag{C2}$$

Thus, one finally obtains for the total Peierls-dimerized spectrum

$$\mathrm{Im}\langle 1|G_{\alpha=1}(z)|1\rangle_D = \pi[1-T^2/T'^2]u(1-T/T')\delta(z) + \frac{1}{8T'^2}\{(4T)^2-[z+4(T^2-T'^2)/z]^2\}^{1/2}, \tag{C3}$$

where $u(1-T/T')$ is the unit step function, and $\delta(z)$ is the Dirac delta function. Of course, for $T=T'$ one reobtains (54).

ACKNOWLEDGMENTS

The author would like to thank Professor J.L. Gammel and Dr. A. Trías for very useful insights, and is indebted to a referee for comments that greatly improved the presentation of the work.

[1]P.A. Fedders, Phys. Rev. B **17**, 40, 2098 (1978).
[2]C.T. Corcoran and P.W. Langhoff, J. Math. Phys. **18**, 651 (1977); J.A. Shohat and J.D. Tamarkin, *The Problem of Moments, Mathematical Surveys* 1 (American Mathematical Society, Providence, R.I., 1950), 2nd ed.
[3]H.S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1948), Chap. V.
[4]M.N. Barber and B.W. Ninham, *Random and Restricted Walks* (Gordon and Breach, New York, 1970), Chap. V.
[5]G.A. Baker, Jr. and J.L. Gammel, Eds., *The Padé Approximant in Theoretical Physics* (Academic, New York, 1970).
[6]G.A. Baker, Jr., *Essentials of Padé Approximants* (Academic, New York, 1975).
[7]P.F. Byrd and M.D. Friedman, *Handbook of Elliptical Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).
[8]S.K. Lyo and J.-P. Gallinar, J. Phys. C **10**, 1693 (1977).
[9]J.-P. Gallinar, Phys. Rev. B **17**, 807 (1978).
[10]D. Kalkstein and P. Soven, Surface Sci. **26**, 85 (1971).
[11]E.A. Di Marzio, J. Chem. Phys. **42**, 2101 (1965).
[12]J.B. Sokoloff, Phys. Rev. B **2**, 3707 (1970).
[13]J.-P. Gallinar, Phys. Rev. B **13**, 4196 (1976).
[14]J. Hubbard, Proc. R. Soc. A **276**, 238 (1963); A **281**, 401 (1964).
[15]G. Beni, T. Holstein, and P. Pincus, Phys. Rev. B **8**, 312 (1973).
[16]F. Oberhettinger, *Tabellen zur Fourier Transformation* (Springer-Verlag, Berlin, 1957).
[17]S.K. Lyo, Phys. Rev. B **14**, 2335 (1976).
[18]S.K. Lyo and T. Holstein, Phys. Rev. B **14**, 5137 (1976).
[19]For $n/N \ll 1$ a formal proof similar to that of Beni *et al.* may be given of this (unpublished).
[20]J.B. Torrance, B.A. Scott, and F.B. Kaufman, Solid State Commun. **17**, 1369 (1975).
[21]This divergence disappears for the even ν crystallographic lines [see Eq. (31)]; which can only be attributed here to the hard core interaction between the spheres.
[22]It should be pointed out that the spectral density with $\alpha=1$ in (44), comes about from the basically one-dimensional reflection of the representative point on lattice plane 2; this resulting from the equal weight ($\alpha=1$) assigned to all lattice points on this plane. As $\nu\to\infty$ this spectral density converges only on the *average* to $[(4T)^2-z^2]^{-1/2}$. Although any measurement of a density of states such as (44), for $\nu\to\infty$, should be indistinguishable from its average, one must notice that the resulting density in (44) cannot be of bounded variation since it has all its moments equal to those of $[(4T)^2-z^2]^{-1/2}$, and yet strikingly differs from it on a set of measure different from zero (see Ref. 3). However, because $\overline{\sin^4\nu\theta}\neq\frac{1}{4}$, a direct measurement of the *square* of the density of states would, so to speak, stabilize the reconstruction of the spectral density from the set of all its moments. The "presence" of the far away barrier would then by this procedure make itself felt.
[23]R.E. Peierls, Quantum Theory of Solids (Oxford U.P., London, 1955), Chap. V.
[24]Y. Lépine and A. Caillé, J. Chem. Phys. **67**, 5598 (1977).
[25]J.-P. Gallinar, Phys. Rev. B **11**, 4421 (1975).
[26]J. Hubbard, Phys. Rev. B **17**, 494 (1978).
[27]W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 1324 (1970).

Potts models, percolation, and duality

J. W. Essam

Department of Mathematics, Westfield College, University of London, Hampstead, London, NW3 7ST. United Kingdom
(Received 4 August 1978)

A λ-state Potts model with multispin interactions is considered which includes models recently introduced by other workers in order to provide a Hamiltonian for site percolation. A duality transformation is obtained for this model which becomes the matching relation for site percolation when $\lambda = 1$.

1. INTRODUCTION

It has been known for some time[1] that the thermodynamic properties of the λ-state Potts model yield corresponding properties of bond percolation in the limite $\lambda \to 1$. Recently Kunz and Wu[2] have extended this result to site percolation by considering a generalization of the Potts model to include multispin interactions. Ashley and Temperley[3] have further developed this idea and shown that different Potts models can yield the same percolation problem.

The duality transformation for the Potts model with two-spin interactions was obtained by Potts[4] using a transfer matrix method. It was later obtained by topological arguments (see for example Refs. 5 and 6) and may be written

$$(z^{*-1} - 1)(z^{-1} - 1) = \lambda, \tag{1.1}$$

where z is defined in terms of the interaction parameter K by $z = e^{-\lambda K}$ and z^* denotes the corresponding variable for the dual problem. In Sec. 3 we generalize Eq. (1.1) to Potts models with multispin interactions. The duality relation for the zero field partition function is seen as an immediate extension of the matching relation for the mean number of clusters in a site percolation problem.[7] The latter is rederived by a simpler method which avoids the use of nonplanar graphs.

In the case $\lambda = 2$, the Potts model with two-spin interactions reduces to an Ising problem with two-spin interactions for which the duality relation was given much earlier by Kramers and Wannier.[8] Duality for Ising models with multispin intractions has recently received considerable attention.[9-13] The $\lambda = 2$ Potts model with three-spin interactions in zero field is also isomorphic with an Ising model with only two-spin interactions. Our result in this case enables the self-duality relation for the Ising model on the triangular lattice to be obtained without the usual reference to the honeycomb lattice (see also Wegner[11]).

In general the $\lambda = 2$ Potts model with l-spin interactions is isomorphic to an Ising model with even interactions of order 2 $[l/2]$ and less, and seems not to have been discussed previously when $l \geqslant 4$ except perhaps as a special case in the articles sited above.[9-13]

The model is introduced in Sec. 2 in a form which includes the conventional Potts model[4] as well as the models of Kunz and Wu[2] and Ashley and Temperley.[3] The correspondence with bond and site percolation is discussed in a unified scheme and the connecting formulas, some of which are required for the main discussion of Sec. 3, are derived.

2. THE MODEL AND ITS RELATION TO PERCOLATION

Consider a system of spins S. Each spin has λ states and selected subsets interact via a Potts interaction. Let the set of all such interacting subsets be denoted by I and define a bipartite interaction graph G which has vertex set $V = (S,I)$ and edge set

$$E = \{(s,i) : s \in S,\ i \in I \text{ and } s \in i.\}.$$

Suppose that each interaction involves at least two spins, so that the vertices of I have at least degree two, and also that spins involved in each interaction $i \in I$ are distinct so that there are no multiedges. Let the states of each spin be indexed by a variable $\alpha = 1, \cdots, \lambda$. In any state of the system let:

$$\xi = \{i \in I : \text{ not all spins of } i \text{ are in the same state}\},$$

$$\eta = \{s \in S : \ s \text{ not in state } \alpha = 1\},$$

$$\zeta = \{i \in I : \text{ not all spins of } i \text{ are in state } \alpha = 1\}.$$

Assign zero energy to the reference state of the system in which all spins are in state $\alpha = 1$ and an additional energy $-kT \log z_i$ for each $i \in \xi$, $-kT \log \mu_s$ for each $s \in \eta$ and $-kT \log \mu_i$ for each $i \in \zeta$. If the variables $\{\mu_s, s \in S\}$, $\{\mu_i, z_i, i \in I\}$ all lie in the interval [0,1], then the reference state is a ground state. The partition function is

$$\Lambda(z,\mu) = \sum_{\text{spin states}} z^{\xi} \mu^{\eta \cup \xi}, \tag{2.1}$$

where

$$z^{\xi} = \prod_{i \in \xi} z_i \text{ and } \mu^{\eta \cup \xi} = \prod_{s \in \eta} \mu_s \prod_{i \in \xi} \mu_i. \tag{2.2}$$

(This convention for a variable raised to the power of a set enables all interaction variables to be considered as distinct without making the notation unduly heavy.) The model so defined will be called a generalized Potts model and denoted by (S,I).

Now consider the percolation model P on G in which each vertex of S is "occupied" with probability 1 and each vertex i of I is occupied independently with probability p_i. A percolation state may be specified by giving the vertices $I' \subseteq I$ which are occupied, and such a state occurs with probability $p^{I'}(1-p)^{I \setminus I'}$. A path is a sequence of vertices, successive members of which constitute an edge of E. An occupied path is one in which all vertices are occupied. In any percolation state two vertices are connected if there is an occupied path between them, and a maximal set of connected vertices is called a cluster. A single spin will be thought of

 0022-2488/79/081769-05$01.00

as connected to itself, and so a cluster may be single isolated spin. The "percolation average" of a random variable $X(I')$ is defined by

$$\langle X\rangle_p = \sum_{I'\subseteq I} p^{I'}(1-p)^{I\setminus I'} X(I'). \tag{2.3}$$

For example $X(I')$ may be the number of clusters or the number of spins connected to a given spin when the vertices I' are occupied.

We now demonstrate that the partition function of (S,I) may be written as a percolation average for P. First note the identities

$$z^{\xi} = \prod_{i\in I}[z_i + (1-z_i)\delta(i)]$$

$$= \sum_{I'\subseteq I} z^{I\setminus I'}(1-z)^{I'}\prod_{i\in I'}\delta(i), \tag{2.4}$$

where the indicater variable $\delta(i)$ is defined by

$$\delta(i) = \begin{cases} 1, & \text{if } i\notin\xi \\ 0, & \text{if } i\in\xi. \end{cases} \tag{2.5}$$

With $p_i = 1 - z_i$ and taking I' to be vertices of I which are occupied in P we may write

$$z^{\xi} = \langle\prod_{i\in I'}\delta(i)\rangle_p. \tag{2.6}$$

By definition of ξ we see that

$$\prod_{i\in I'}\delta(i) = \begin{cases} 1, & \text{if all spins in each cluster are} \\ & \text{in the same state,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

Substituting (2.6) into (2.1) and interchanging the sum and average we obtain,

$$\Lambda(1-p,\mu) = \left\langle \prod_{c\in\mathscr{C}'}[1+(\lambda-1)\mu^{V(c)}] \right\rangle_p, \tag{2.8}$$

where $\mathscr{C}'$ is the set of all finite clusters corresponding to I' and $V(c)$ is the subset of vertices V which belong to the cluster c. In restricting the product to finite clusters we are assuming that sufficient μ variables are less than one, so that $\mu^{V(c)} = 0$ for any infinite cluster.

The result (2.8) will now be used to relate the thermodynamic properties of (S,I) to the percolation functions of P. We shall take $\mu = 1$ to mean $\mu_v \to 1^-, \forall\, v\in V$.

The free energy $\ln\Lambda$ yields a generating function $G(p,\mu)$ for percolation theory, since

$$\mathscr{G}(p,\mu) \equiv \frac{\partial}{\partial\lambda}\ln\Lambda\Big|_{\lambda=1} = \left\langle \sum_{c\in\mathscr{C}'}\mu^{V(c)} \right\rangle_p. \tag{2.9}$$

When $\mu = 1$, $\mathscr{G}(p,u) = \langle n\rangle_p$, the mean number of finite spin clusters, which for an infinite G is normally calculated per spin or per interaction, since $\mathscr{G}$ itself would be infinite.

Differentiating (2.9) w.r.t. μ_v,

$$\frac{\partial\mathscr{G}}{\partial\ln\mu_v} = \langle\delta_v\mu^{V(c_v)}\rangle_p, \tag{2.10}$$

where $\delta_v = 1$ if there is a finite cluster c_v containing vertex v but zero otherwise. Setting $\mu = 1$ gives the probability that v is part of a finite cluster, so that if G is finite,

$$\frac{\partial G}{\partial\mu_v}\Big|_{\mu=1} = \begin{cases} 1, & v\in S, \\ p_v, & v\in I. \end{cases} \tag{2.11}$$

For G infinite and $v\in V$ the percolation probability P_v may be defined as the conditional probability that, given v belongs to a cluster, it belongs to an infinite cluster, and hence,

$$P_v = 1 - p_v^{-1}\frac{\partial\mathscr{G}}{\partial\mu_v}\Big|_{\mu=1}, \tag{2.12}$$

where $p_v = 1$ if $s\in S$.

Setting $\mu_s = 1$, $\forall\, s$ and $\mu_i = \mu, \forall\, i$ in (2.10) gives

$$\frac{\partial\mathscr{G}}{\partial\ln\mu_v}\Big|_{\substack{\mu_s=1\\ \mu_v=\mu}} = \langle\delta_v\mu^{|I(c_v)|}\rangle, \tag{2.13}$$

where $I(c)$ is the set of interaction vertices in c. $|I(c_v)|$ is a measure of the cluster size, and if $v\in I$, it is the standard measure of cluster size in the site percolation problem $\bar{P}$, defined later. Equation (2.13) therefore gives the moment generating function for the cluster size distribution. The moment generating function for $|S(c_v)|$ may be obtained by setting $\mu_s = \mu$, $\forall s$ and $\mu_i = 1$, $\forall\, i$. This is an alternative measure of the cluster size and is, in fact, the number of sites in a bond problem (see later), whereas $|I(c_v)|$ is the number of bonds (see also Stephen[14]).

For $v, v'\in V$, the pair connectedness $P_{vv'}$ is the probability that v and v' belong to the same finite cluster. This may be obtained from (2.10) by differentiating first with respect to $\mu_{v'}$ and then with respect to μ_v, since the only terms which survive the second differentiation are the ones where v and v' belong to the same member of $\mathscr{C}'$, thus

$$P_{vv'} = \frac{\partial^2\mathscr{G}}{\partial\ln\mu_v\,\partial\ln\mu_{v'}}\Big|_{\mu=1},$$

$$= \frac{\partial^2\mathscr{G}}{\partial\mu_v\,\partial\mu_{v'}}\Big|_{\mu=1} \quad \text{if } v\neq v'. \tag{2.14}$$

The corresponding derivative of $\ln\Lambda$ is a spin correlation function.

The above relationship between percolation and the Potts model has been derived by other authors for less general models.

Kastelyn and Fortuin[1] demonstrated a correspondence between Potts models with only two-spin interactions and bond percolation. This was recently developed by Stephen[14] and Wu.[15] For such a model (S,I) the vertices I of the graph G have degree two, and for each $i\in I$ the pair of edges $\{(s_1,i),(i,s_2)\}$ may be replaced by a single edge (s_1,s_2) to give a graph H. This operation is called suppression of the vertices I. The edges so formed are deemed to be occupied when the corresponding vertex of i is occupied and an occupied edge provides a connection between its terminal vertices. The percolation model P corresponding to (S,I) is clearly the bond problem on H. Figure 1(a) illustrates the Potts model for the bond problem on the square lattice.

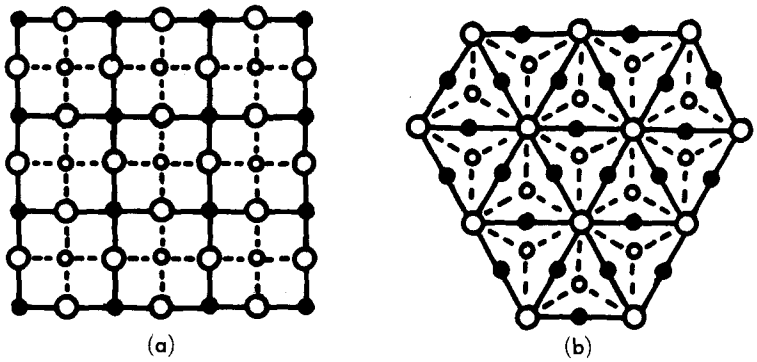

FIG. 1. (a) Kasteleyn–Fortuin correspondence between a Potts model with two-spin interactions and the bond problem on the square lattice. The self duality of this model is illustrated (b) Kunz–Wu correspondence between a Potts model with six-spin interactions and the site problem on the triangular lattice. The dual model is also shown.

Key : 0 = interaction vertex, • = primary spin, ° = dual spin,— = primary edge, ---- = dual edge.

The percolation model P corresponding to any Potts model (S,I) may be converted into a more conventional percolation model $\bar{P}$ in which all sites are randomly occupied. The sites S of G which are occupied with probability one may be removed by the following "star" transformation. For each vertex $s \in S$ let $I(s)$ be the vertices of I which are adjacent to s. Then replace the edges $\{(s,i), i \in I(s)\}$ by the complete graph on $I(s)$ and remove s from the vertex set of G. If s has degree two this corresponds to suppression. If s has degree three it is the star–triangle transformation. Multiedges formed during the transformation may be replaced by single edges without changing the connectedness of the model. The resulting graph will be called $\bar{G}$. The site clusters of $\bar{P}$ correspond to the clusters of P, but a cluster of P which is just a single spin has no counterpart in $\bar{P}$. However, the following simple relation holds between the mean number of clusters for the two problems

$$\langle n \rangle_p = \langle \bar{n} \rangle_p + \sum_{s \in S} (1-p)^{I(s)}. \tag{2.15}$$

The pair connectedness for two vertices $i_1, i_2 \in I$ is the same in both problems.

A given site problem may correspond to several Potts models. Kunz and Wu[2] showed that the site problem on a graph H corresponds to the Potts model (S,I) obtained by taking I to be the sites of H and associating a vertex of S with each edge of H. Clearly H is the graph $\bar{G}$ defined above. The Potts model for the site problem on the triangular lattice is shown in Fig. 1(b). This model involves six-spin interactions. An alternative Potts model for which $\bar{P}$ is the site problem on the triangular lattice was given by Ashley and Temperley.[3] This model is obtained by taking S and I to be the two triangular sublattices of a honeycomb lattice as shown in Fig. 2. In contrast to the Kunz-Wu model only three-spin interactions are involved.

It is well known that any bond problem may be transformed into a site problem on a different graph. In the present context, if the bond problem on a graph H corresponds by the Kastelyn-Fortuin transformation[1] to a Potts model (S,I), then the equivalent site problem is $\bar{P}$ described above.

3. DUALITY FOR PERCOLATION AND THE ZERO FIELD PARTITION FUNCTION

Consider a Potts model (S,I) and for simplicity assume that all vertices have at least degree two. Assume also that the interaction graph G is connected and planar, and that it has been drawn in the plane to form a plane graph (also denoted by G). The graph $\bar{G}$ obtained from G by the star transformation is not necessarily planar, and is in general a decorated mosaic.[7] The dual model (S^*,I) is constructed by placing a vertex of S^* in each face of G and connecting it by an edge to each vertex of I which lies in the boundary of the face in such a way that no two edges intersect. The plane graph so formed will be denoted by G^* and $G^{**} = G$. Figure

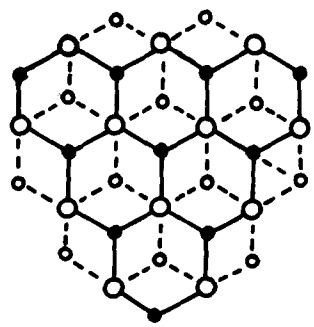

FIG. 2. Ashley–Temperley correspondence between a Potts model with three-spin interactions and the site problem on the triangular lattice. The key is as in Fig. 1.

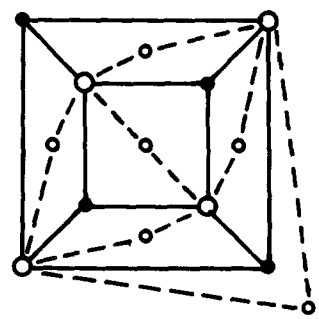

FIG. 3. Duality on a finite graph. Both Potts models correspond to the site problem on the tetrahedron. The key is as in Fig. 1.

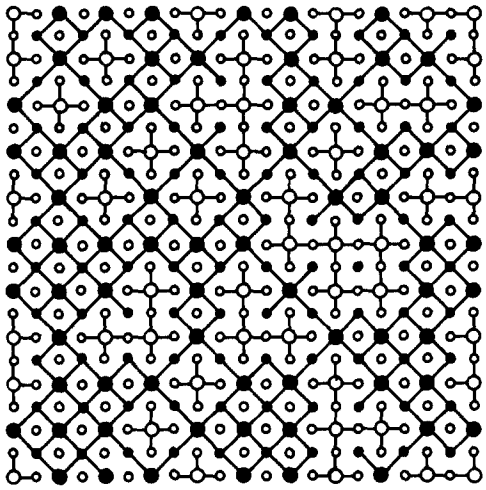

FIG. 4. Corresponding percolation states for the site problem on the square lattice and its matching lattice. Open circles are occupied on the square lattice, whereas solid circles are occupied on the matching lattice.

3 shows the dual transformation between the cube G and the decorated tetrahedron G^*. Notice that the Ashley–Temperley model of Fig. 2 is self-dual, as is the square lattice with two-spin interactions [Fig. 1(a)]. The dual of the Kunz–Wu model of Fig. 1(b) is the diced lattice in which the sixth order vertices are interaction vertices. The percolation model $\bar{P}$ corresponding to the latter model is also the triangular site problem.

The percolation model $\bar{P}^*$ corresponding to (S^*,I) together with $\bar{P}$ constitute a pair of matching site problems.[7] These have the property that the mean number of clusters at probability p for $\bar{P}$ is related to the mean number of clusters at probability $1-p$ for $\bar{P}^*$. (For the models in Fig. 3, $\bar{P}$ and $\bar{P}^*$ are both the site problem on the tetrahedron which is self-matching.) We now obtain a similar relation for P and P^* which by (2.17) implies the matching relation between $\bar{P}$ and $\bar{P}^*$. A relation between the partition functions of dual models is also obtained by the same technique.

Suppose that the occupied vertices of G in a percolation state of P are colored black. This means that all the S vertices are black, together with the subset I' of I. The black vertices together with the edges of G connecting black vertices form a subgraph which we shall call G_B, the components of which are the clusters. A corresponding state of P^* may be obtained by supposing that the vertices $I \setminus I'$ are occupied on G^*. If these vertices together with the vertices S^* are colored white, the white vertices define a subgraph G_W the components of which are the clusters for this state of P^*. Figure 4 shows a pair of dual models and the corresponding states. The model (S^*,I) is the Kunz–Wu model[2] for which $\bar{P}^*$ is the site problem on the square lattice, and its dual (S,I) is such that $\bar{P}$ is the site problem on the square lattice with first and second-neighbor connections. The latter is the matching lattice of the square lattice. The figure illustrates the following theorem.

Theorem: With G_B and G_W defined as above:

(a) every component of G_W is contained within a single face of G_B, and

(b) each face of G_B contains exactly one component of G_W. The same statement is true with G_B and G_W interchanged by symmetry of the dual construction.

Proof: G_B and G_W never intersect, since the only possible intersection points are vertices of I, and in any percolation state each vertex of I either belongs to G_B or to G_W. It follows that any two vertices which are connected in G_W must lie in the same face of G_B, which proves part (a).

To prove part(b) we first notice that the faces of G_B are made up of faces of G each of which contains a vertex of S^*. It follows immediately that every face of G_B contains at least one component of G_W. To show that there cannot be more than one, we must prove that any two vertices of G_W which lie in the same face of G_B are connected on G_W. It is sufficient to prove this for two vertices of S^*, since any vertex of $I \setminus I'$ is on the boundary of some face of G and is therefore connected to a vertex of S^*. Suppose that the vertices s_1 and s_2 of S^* lie in the faces F_1 and F_2 of G which are within the face F_B of G_B. Since F_B is a connected region of the plane, the faces F_1 and F_2 must be connected by a chain of faces of G which are within F_B, successive members of which have edge to edge contact along at least one edge of G which is internal to F_B. An edge of G which is internal to F_B must have a white vertex which is in $I \setminus I'$, and hence vertices of S^* which lie in successive faces of the chain must be connected on G_W via this white vertex. Hence the vertices s_1 and s_2 are connected by a chain of white vertices.

It follows from the theorem that the number of faces f_B of G_B, including the infinite face, is equal to the number of components n_W of G_W, which is the number of clusters for the problem P^*. Also the number of faces of G_B is determined in terms of n_B by Euler's law applied to G_B, thus,

$$n_W = f_B = e_B - v_B + n_B + 1. \tag{3.1}$$

On taking averages over all percolation states,

$$\langle n_B \rangle_p = \Phi(p) + \langle n_W \rangle^*_{1-p}, \tag{3.2}$$

where the average on the right is calculated for P^*, for which $1-p_i$ is the probability of occupation of i. The function $\Phi(p)$ is linear in the variables p_i, and is given by

$$\Phi(p) = |S| + \sum_{i\in I} p_i - \sum_{i\in I} \gamma_i p_i - 1, \tag{3.3}$$

where γ_i is the number of vertices of S which are adjacent to i. Notice that

$$\begin{aligned}\Phi(1-p) &= |S| + |I| - |E| - \sum_{i\in I} p_i + \sum_{i\in I} \gamma_i p_i - 1,\\ &= -|S^*| - \sum_{i\in I} p_i + \sum_{i\in I} \gamma_i p_i + 1,\\ &= -\Phi^*(p),\end{aligned} \tag{3.4}$$

where we have used (3.3) with $p_i = 1$. Notice also that by construction γ_i is also the number of vertices of S^* adjacent to i.

The matching relation obtained by Sykes and Essam[7] was

$$\langle \bar{n}_B \rangle_p = N\phi(p) + \langle \bar{n}_W \rangle^*_{1-p}, \tag{3.5}$$

where $N = |I|$. Using (2.7) we find

$$N\phi(p) = \Phi(p) + \sum_{s \in S^*} p^{I(s)} - \sum_{s \in S} (1-p)^{I(s)}, \tag{3.6}$$

which should be compared with their equation (6.14).

The above relations are exact when G is a finite graph but should also apply to an infinite lattice graph in the thermodynamic limit, in which case one works with the mean number of clusters per vertex of I. A proper treatment of this limit will not be attempted here. However, we note that although our main theorem is true when the lattice is made finite by embedding it in a torus in the usual way (see Fig. 4), not all subgraphs G_B are properly embedded in the torus. This creates difficulties when applying Euler's formula for the number of regions.

As an illustration of the above formulas if P is the site problem on the square lattices with first and second neighbor connections,

$$\Phi(p) = N(1-3p)$$

and

$$\phi(p) = 1 - 3p + 2p^2 - (1-p)^4,$$

$$= p - 4p^2 + 4p^3 - p^4,$$

in agreement with Table II of Sykes and Essam.[7]

A duality relation for the zero field partition function will now be obtained. Setting $\mu = 1$ in (2.8),

$$\Lambda(1-p,1) = \langle \lambda^n \rangle_p, \tag{3.7}$$

where $n = |\mathscr{C}'|$, the number of finite clusters in the percolation model P. The duality for Λ may therefore be considered as a percolation problem, and the notation of the first part of the section will be maintained. By definition,

$$\langle \lambda^n \rangle_p \equiv \langle \lambda^{n_B} \rangle_p = \sum_{I' \subseteq I} p^{I'} (1-p)^{I \setminus I'} \lambda^{n_B}. \tag{3.8}$$

From (3.1)

$$n_B = n_W + |S| + |I'| - \sum_{i \in I'} \gamma_i - 1, \tag{3.9}$$

which is Eq. (3.3) before averaging. Substituting in (3.8) gives

$$\langle \lambda^{n_B} \rangle_p = \lambda^{|S|-1} \sum_{I' \subseteq I} (p\lambda^{1-\gamma})^{I'} (1-p)^{I \setminus I'} \lambda^{n_W}, \tag{3.10}$$

and renormalizing the distribution gives

$$\langle \lambda^{n_B} \rangle_p$$

$$= \lambda^{|S|-1}(1-p+p\lambda^{1-\gamma})^I \sum_{I' \subseteq I} p^{*I'} (1-p^*)^{I \setminus I'} \lambda^{n_W},$$

$$= \lambda^{|S|-1}(1-p+p\lambda^{1-\gamma})^I \langle \lambda^{n_W} \rangle^*_{p^*}, \tag{3.11}$$

where

$$p^* = \frac{1-p}{1-p+p\lambda^{1-\gamma}}. \tag{3.12}$$

When $\lambda = 1, p^* = 1-p$, which is the usual percolation relation. In terms of the Potts model variables,

$$\Lambda(z,1) = \lambda^{|S|-1}(z + (1-z)\lambda^{1-\gamma})^I \Lambda^*(z^*,1), \tag{3.13}$$

where

$$z^* = \frac{(1-z)}{z\lambda^{\gamma-1} + (1-z)}, \tag{3.14}$$

which may be written in the symmetric form

$$(z^{*-1} - 1)(z^{-1} - 1) = \lambda^{\gamma-1}. \tag{3.15}$$

Equations (3.12), (3.14), and (3.15) are local relations and the subscript i on the variables z, p, and γ has been suppressed for convenience. For model with only two-spin interactions $\gamma_i = 2$, $\forall\ i$ and Eq. (1.1) is obtained.

It is worth noting that when $\lambda = 2$ the Ashley–Temperley model[3] is isomorphic to a triangular lattice Ising model with a two-spin interaction parameter $y = z^{1/2}$ for which the duality relation reads (with $\gamma = 3$),

$$(y^{*-2} - 1)(y^{-2} - 1) = 4,$$

and since the model is self-dual, the critical point (assumed to be unique) is located at $y_c^2 = \frac{1}{3}$ as found by Onsager.[16]

ACKNOWLEDGMENT

The author is grateful for valuable discussions with Dr. C.M. Place.

[1]P. Kasteleyn and C. Fortuin, J. Phys. Soc. Jap. (suppl.) **26**, 11–14 (1969).
[2]H. Kunz and F.Y. Wu, J. Phys. C **11**, L1–4 (1978).
[3]S.E. Ashley and H.N.V. Temperley (preprint).
[4]R.B. Potts, Proc. Cambridge. Philos. Soc. **48**, 106 (1952).
[5]J.W. Essam, Discrete Math. **1**, 83–112 (1971).
[6]L. Mittag and M.J. Stephen, J. Math. Phys. **12**, 441–50 (1971). (A topological derivation of duality for a Potts model with two-spin interactions by M.E. Fisher is reported as a private communication in this paper.)
[7]M.F. Sykes and J.W. Essam, J. Math. Phys. **5**, 1117–27 (1964).
[8]H.A. Kramers and G.H. Wannier, Phys. Rev. **60**, 252 (1941).
[9]D. Merlini and C. Gruber, J. Math. Phys. **13**, 1814–23 (1971).
[10]F.J. Wegner, J. Math. Phys. **12**, 2259–72 (1971).
[11]F.J. Wegner, Physica **68**, 570–8 (1973).
[12]D.W. Wood, J. Phys. C **5**, L181–3 (1972).
[13]D.W. Wood and N.E. Pegg, J. Phys. A **10**, 229–38 (1977).
[14]M.J. Stephen, Phys. Rev. **15**, 5674–9 (1977).
[15]F.Y. Wu, J. Stat. Phys. **18**, 115–23 (1978).
[16]L. Onsager, Phys. Rev. **65**, 117 (1944).

Husimi–Temperley model under a random field

Tsuyoshi Horiguchi

Department of Applied Science, Faculty of Engineering, Tohoku University, Sendai 980, Japan
(Received 14 November 1978; revised manuscript received 6 February 1979)

The free energy of the Husimi–Temperley model under a random field is obtained in terms of the n-replica method. The interchangeability of various limits such as the thermodynamic limit, the limit $n\to 0$ where n is the number of replicas, etc., is justified on the usage of the n-replica method in the course of the evaluation of the free energy. A tricritical point is found for the model under the random field which takes on two values σ and $-\sigma$ with equal probability.

The n-replica method has been used much in the course of the evaluation of the free energy of random quenched systems and has contributed to the development of the spin-glass theory.[1-3] However, this method contains several points to be clarified on its usage, e.g., the existence of the thermodynamic limit, the order of the $n\to 0$ and the thermodynamic limit,[4,5] the analytic continuation of the function of integer n to the one of nonintegral n,[6,7] the minimization of the free energy by the method of steepest descent,[5] etc.

The Husimi–Temperley model (or Weiss type model)[8] under the random field distributed according to a Gaussian distribution has been discussed by Schneider and Pytte[9] as a simple example to be solved by the n-replica method. They have shown that their results obtained by the formal usage of the n-replica method are exact by constructing tight upper and lower bounds for the free energy by using the Gibbs–Bogoliubov inequality.

It is, however, interesting to clarify and justify some mathematical steps such as interchangeability of various limits, etc. mentioned above on the direct usage of the n-replica method in a calculation of the free energy. These troublesome steps are avoided or ignored in the previous calculations.[7,9,10] In this note, we discuss the Husimi–Temperley model under the random field distributed according to the various types of distrubution in a different viewpoint of the n-replica method from the one described previously in order to clarify and justify the points mentioned above.

The Hamiltonian considered here is given by

$$\mathcal{H}_N = -\frac{J}{N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ i\neq j}}^{N} s_i s_j - \sum_{i=1}^{N} h_i s_i, \quad s_i = \pm 1, \tag{1}$$

where $J>0$ and h_i is the random field distributed according to $p(h_i)$. $p(h_i)$ takes the Gaussian distribution $N(\bar{h},\sigma^2)$, the uniform distribution $U(\bar{h}-3^{1/2}\sigma,\bar{h}+3^{1/2}\sigma)$ or the one expressed by

$$p(h_i) = \sum_{j=1}^{l} c_j \delta(h_i - \tilde{h}_j) \tag{2}$$

with $\Sigma_{j=1}^{l} c_j = 1$, $\Sigma_{j=1}^{l} c_j \tilde{h}_j = \bar{h}$, and $\Sigma_{j=1}^{l} c_j \tilde{h}_j^2 - \bar{h}^2 = \sigma^2$. When we consider the site-random Husimi–Temperley model under the uniform field, the Hamiltonian is immediately reduced to Eq. (1) with

$$p(h_i) = c\delta(h_i - \sigma) + (1-c)\delta(h_i + \sigma). \tag{3}$$

Our aim is to find the free energy per site when the number of the total sites N goes to infinity:

$$-\beta f = \lim_{N\to\infty} \frac{1}{N} \langle \log Z_N(\{h_i\})\rangle_N, \tag{4}$$

where $\beta = 1/k_B T$ as usual and

$$Z_N(\{h_i\}) = \sum_{s_1=\pm 1}\cdots\sum_{s_N=\pm 1} e^{-\beta\mathcal{H}_N} = 2^N e^{-\beta J}\left(\frac{N}{\pi}\right)^{1/2} \times \int_{-\infty}^{\infty} dt e^{-Nt^2} \prod_{i=1}^{N} \cosh[2(\beta J)^{1/2}t + \beta h_i] \tag{5}$$

and

$$\langle Q_N(\{h_i\})\rangle_N = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \prod_{i=1}^{N} [dh_i p(h_i)] Q_N(\{h_i\}). \tag{6}$$

Equation (4) is expressed as

$$-\beta f = \lim_{N\to\infty}\lim_{x\to+0} \frac{d}{dx}\Phi_N(x), \tag{7}$$

where

$$\Phi_N(x) = \langle Z_N^x(\{h_i\})\rangle_N^{1/N}. \tag{8}$$

We note that $\Phi_N(z)$ is an analytic function of z in the domain $D=\{z \mid \mathrm{Re}z \geqslant -\epsilon\}$, where ϵ is a small positive number, when we choose $\arg Z_N^z(\{h_i\}) = 0$ and $\arg\langle Z^z(\{h_i\})\rangle_N^{1/N} = 0$ when $z = x > 0$. We shall later show that $\{\Phi_N(z)\}$ converges uniformly as $N\to\infty$ in D. Then we have

$$-\beta f = \lim_{x\to+0} \frac{d}{dx}\phi(x), \tag{9}$$

where

$$\phi(x) = \lim_{N\to\infty} \Phi_N(x). \tag{10}$$

It is easy to show that $\Phi_N(x)$ is bounded. Then $\{\Phi_N(x)\}$ has at least one limit. However we could not show directly that $\{\Phi_N(x)\}$ has a unique limit. For instance, we have

$$\Phi_N(x) = \left\langle\left(\sum_{l=1}^{M} A_N^{-1/2}(t_l) e^{-NF_N(t_l)}\right)^x\right\rangle_N^{1/N}\left\{1 + O\left(\frac{x}{N^2}\right)\right\} \tag{11}$$

where

$$F_N(t) = t^2 - \frac{1}{N}\sum_{i=1}^{N} \log\cosh[2(\beta J)^{1/2}t + \beta h_i], \tag{12}$$

 0022-2488/79/081774-02$01.00

$$A_N(t) = 1 - \frac{2\beta J}{N} \sum_{i=1}^{N} \operatorname{sech}^2[2(\beta J)^{1/2}t + \beta h_i]. \tag{13}$$

$F_N(t)$ has at least one minimum and at most $N+1$ minima which are assumed to be attained at t_l, where $l = 1,2,...,M$ with $1 \leqslant M \leqslant N+1$, and are denoted $F_N(t_1) = F_N(t_2) = \cdots = F_N(t_M)$. $A_N(t_l)$ is positive and bounded. t_l is a function of $\{h_1,h_2,...,h_N\}$ and determined by the following equation:

$$t_l - \frac{(\beta J)^{1/2}}{N} \sum_{i=1}^{N} \tanh[2(\beta J)^{1/2}t_l + \beta h_i] = 0. \tag{14}$$

One of the annoying problems is the convergency of t_l as $N \to \infty$.

By considering n replicas of the system, we calculate $\Phi_N(n)$ for positive integers n and obtain

$$\psi(n) = \lim_{N\to\infty} \Phi_N(n) = \max_{-\infty<t<\infty} \{e^{-nt^2} 2^n \langle \cosh^n[2(\beta J)^{1/2}t + \beta h]\rangle\}, \tag{15}$$

where $\langle Q(h)\rangle$ denotes the average with $N=1$ defined by (6). As mentioned previously,[6,7] $\psi(x)$ for real $x>0$ is in general different from $\phi(x)$.

Now let us prove that $\phi(x) = \psi(x)$ for $x>0$, though it is clear for $x=0$. First we find the sequence of the functions $\{\Psi_N(x)\}$ for $x>0$ which converges to $\psi(x)$ as $N\to\infty$;

$$\Psi_N(x) = \left\langle \int_{-\infty}^{\infty} dt\, e^{-xNt^2} \prod_{i=1}^{N} \{2^x \cosh^x[2(\beta J)^{1/2}t + \beta h_i]\} \right\rangle_N^{1/N} \tag{16}$$

$$= \left\{ \int_{-\infty}^{\infty} dt\, e^{-xNt^2} 2^{xN} \langle \cosh^x[2(\beta J)^{1/2}t + \beta h]\rangle^N \right\}^{1/N}. \tag{17}$$

Consider the sequence $\{\tilde{\Psi}_N(x)\}$ defined as follows:

$$\tilde{\Psi}_N(x) = \Psi_{N+1}^{N+1}(x)/\Psi_N^N(x). \tag{18}$$

We have

$$\tilde{\Psi}_N(x) \leqslant e^{x\beta J} 2^x \langle \cosh^x(\beta h)\rangle, \tag{19}$$

and by the Schwarz's inequality applied to (17)

$$\tilde{\Psi}_N(x) \leqslant \tilde{\Psi}_{N+1}(x). \tag{20}$$

Then $\{\tilde{\Psi}_N(x)\}$ and $\{\Psi_N(x)\}$ converge to a unique limit which is expressed in the following two ways by using expressions (16) and (17), respectively:

$$\psi(x) = \lim_{N\to\infty} \left\langle \sum_{l=1}^{M} \left(\frac{1}{\pi} xNA_N(t_l)\right)^{-1/2} e^{-xNF_N(t_l)} \right\rangle_N^{1/N} \times [1 + O(1/N^2x)] \tag{21}$$

$$= \max_{-\infty<t<\infty} \{e^{-xt^2} 2^x \langle \cosh^x[2(\beta J)^{1/2}t + \beta h]\rangle\}. \tag{22}$$

where $F_N(t)$ and $A_N(t)$ are given by (12) and (13). t_l satisfies Eq. (14). Because $A_N(t_l)$ is positive and bounded, we have from (10), (11), (21), and (22)

$$\phi(x) = \psi(x) = e^{-xt_0^2} 2^x \langle \cosh^x[2(\beta J)^{1/2}t_0 + \beta h]\rangle, \tag{23}$$

where t_0 is the value which maximizes the expression (22) and satisfies the following equation:

$$t_0 = (\beta J)^{1/2} \langle \cosh^x[2(\beta J)^{1/2}t_0 + \beta h] \tanh[2(\beta J)^{1/2}t_0 + \beta h]\rangle / \langle \cosh^x[2(\beta J)^{1/2}t_0 + \beta h]\rangle. \tag{24}$$

Thus we have shown above that $\{\Phi_N(x)\}$ for $x \geqslant 0$ converges to the unique limit. This is equivalent to the existence of the thermodynamic limit. By virtue of the Vitali's theorem, $\{\Phi_N(z)\}$ converges uniformly to the limit $\phi(z)$ as $N\to\infty$ for all $z \in D$; hence $\phi(z)$ is an analytic function of z for all $z \in D$. This fact warrants the transformation of Eq. (7) to Eq. (9). Thus we obtain the free energy as follows:

$$-\beta f = -\beta J m_1^2 + \log 2 + \langle \log \cosh(2\beta J m_1 + \beta h)\rangle, \tag{25}$$

where m_1 satisfies the equation

$$m_1 = \langle \tanh(2\beta J m_1 + \beta h)\rangle. \tag{26}$$

m_1 is the configurational average of the magnetization. The Edwards–Anderson order parameter m_2 is expressed by

$$m_2 = \langle \tanh^2(2\beta J m_1 + \beta h)\rangle. \tag{27}$$

According to the method described in this note, the obtained results are also correct for the random field which distribution assures the uniform convergence of the integral $\langle Z_N^z(\{h_i\})\rangle_N$ for z in any finite region contained in D.

Finally we shall discuss the phase boundary. When $p(h)$ takes the uniform distribution $U(-3^{1/2}\sigma, 3^{1/2}\sigma)$, the obtained phase diagram is similar to the one for the Gaussian distribution $N(0,\sigma^2)$ given by Schneider and Pytte.[9] For $2J/\sigma < 3^{1/2}$, there is no ferromagnetic phase. The phase boundary of the second order phase transition between the phase with $m_1 = 0$ and $m_2 \neq 0$ and the one with $m_1 \neq 0$ and $m_2 \neq 0$ is given by

$$2J/\sigma = 3^{1/2} \coth(3^{1/2}\beta\sigma). \tag{28}$$

On the other hand, when $p(h)$ takes the form given by (3) with $c = \frac{1}{2}$, the phase bounday of the second order phase transition is given by

$$2J/\sigma = \cosh^2\beta\sigma/\beta\sigma, \tag{29}$$

up to the point $2J/\sigma = 3/2\cosh^{-1}(\frac{3}{2})^{1/2} = 2.278$ and $k_BT/\sigma = 1/\cosh^{-1}(\frac{3}{2})^{1/2} = 1.519$ which is the tricritical point. Below this point, we have the phase boundary of the first order phase transition which is determined by Eqs. (25) and (26) up to the point $2J/\sigma = 2$ and $k_BT/\sigma = 0$. For $2J/\sigma < 2$, there is no ferromagnetic phase.

ACKNOWLEDGMENTS

The author wishes to thank Professor T. Morita for numerous helpful discussions and Professor S. Katsura and Professor K. Hiroike for valuable comments.

[1] S.F. Edwards and P.W. Anderson, J. Phys. F **5**, 965 (1975).
[2] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).
[3] A. Aharony, J. Phys. C **11**, L457 (1978).
[4] D.J. Thouless, P.W. Anderson, and R.G. Palmer, Phil. Mag. **35**, 593 (1977).
[5] J.R.L. de Almeida and D.J. Thouless, J. Phys. A **11**, 983 (1978).
[6] T. Morita and T. Horiguchi, Solid St. Commun. **19**, 833 (1976).
[7] M.W. Klein, J. Phys. F **7**, L267 (1977).
[8] S. Katsura, Advan. Phys. **12**, 391 (1963).
[9] T. Schneider and E. Pytte, Phys. Rev. B **15**, 1519 (1977).
[10] D. Sherrington, Phys. Lett. A **58**, 36 (1976).

The spectrum of the multigroup neutron transport operator for bounded spatial domains[a)]

Edward W. Larsen
Theoretical Division, University of California, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

(Received 5 January 1979)

The spectrum of the multigroup neutron transport operator A is studied for bounded spatial regions D which consist of a finite number of material subregions. Our main results provide simple conditions on the material cross sections which guarantee that (1) A possesses eigenvalues in the finite plane; (2) A possesses a "leading" eigenvalue λ_0 which is real, not less than the real part of any other eigenvalue, and to which there corresponds at least one nonnegative eigenfunction $\psi_{\lambda 0}$; and (3) A possesses a "dominant" eigenvalue λ_0 which is real, simple, greater than the real part of any other eigenvalue, and whose eigenfunction $\psi_{\lambda 0}$ satisfies $\psi_{\lambda 0} \geq 0$ and $\int \psi_{\lambda 0} d^2\Omega > 0$. We give examples to illustrate the results and to show that a leading eigenvalue need not be simple, nor its eigenfunction(s) positive.

I. INTRODUCTION

The problem of determining properties of the spectrum of the neutron transport operator has been considered by many authors under varying assumptions on the type of medium and the scattering process.[1-7] Nevertheless, certain gaps have persisted, in particular regarding the existence and multiplicity of an eigenvalue with greatest real part and the positivity of its eigenfunction(s).

The purpose of this article is to combine some results of Jörgens,[2] some more recent developments of Angelescu and Protopopescu,[7] and some new analysis to provide a fairly complete qualitative description of the spectrum of the multigroup transport operator A in a bounded, convex, physical domain D.

The following description, due to Jörgens,[2] provides the starting point of our analysis. If $\sigma[A]$, the spectrum of A, contains points in the finite part of the complex plane, then they are discrete and situated in a left half space, finite in any strip $\beta_1 \leqslant \mathrm{Re}\lambda \leqslant \beta_2$, and the generalized eigenspaces are all finite-dimensional.

To describe our new results, we must make two definitions. First, let a "leading" eigenvalue of A be one which is real, which is not less than the real part of any other eigenvalue, and to which corresponds at least one nonnegative eigenfunction. Next, let a "dominant" eigenvalue of A be one which is real, which is strictly greater than the real part of any other eigenvalue, which is simple, and to which corresponds a nonnegative eigenfunction $\psi(\mathbf{x},\mathbf{\Omega})$ such that $\int \psi(\mathbf{x},\mathbf{\Omega}) d^2\Omega$ is everywhere positive in D.

Then our main results, stated in physical terms, are as follows:

(I) $\sigma[A]$ is nonempty if there exists at least one energy group with self-scattering in some nonempty sphere $|\mathbf{x} - \mathbf{x}_0| < r_0$ in D.

(II) A leading eigenvalue of A exists if and only if $\sigma[A]$ is nonempty.

(III) A dominant eigenvalue of A exists if the condition stated in (I) is met *and* if any neutron, occurring at any point in D and in any energy group, has a positive probability of (eventually) being scattered into any other energy group.

We also give examples to illustrate that a leading eigenvalue need not be simple, nor its eigenfunction(s) positive.

The specific problem of examining the leading eigenvalue of a transport operator has been considered previously by Vidav[5] and Angelescu and Protopopescu[7] for the case of a continuous energy variable which tends to zero and a strictly positive kernel in the scattering integral. (Since the energy variable tends to zero, the spectrum of the transport operator contains a half-space,[4-7] $\mathrm{Re}\lambda \leqslant \lambda^*$, which is often called "the continuum.") Vidav's analysis, which can be directly applied to the multigroup transport operator, shows that if there exists an isolated point eigenvalue in the spectrum of the transport operator (i.e., out of the continuum), then there exists a leading eigenvalue λ_0 whose geometrical multiplicity is 1 and whose (single) eigenfunction is strictly positive in the interior of D. Angelescu and Protopopescu[7] have shown recently that for this problem, λ_0 is in fact dominant. (Thus λ_0 is simple and greater than the real part of any other eigenvalue.) However, for technical reasons, this analysis cannot be directly applied to the multigroup transport operator. [These authors[7] use the continuity of the energy variable to construct an integral operator which is (required to be) completely continuous. For discrete energies, this construction cannot be done.] In any case, our results are more general because we do not require the kernel in the scattering integral (which, for multigroup problems, is a matrix) to be strictly positive.

In this article we formulate conditions on the multigroup transport operator which specifically apply to the class of problems for which a numerical calculation of the eigenvalues (or, in particular, of the dominant eigenvalue) is likely to be attempted. We do this because the multigroup approximation is made basically for the purpose of comput-

a)Work performed under the auspices of the U.S. Department of Energy.

 0022-2488/79/081776-07$01.00

ing, and so there is little reason to allow more general types of behavior in our analysis than is possible in a computer code. Thus we require all of the cross sections to be piecewise continuous (rather than, for example, square-integrable) and the boundaries of the physical system and all of its material subregions to be continuous. However, in order to apply Jörgens' results directly, i.e., without introducing extensions of his theory, we allow the fluxes to be square-integrable. Then the fluxes *can* be piecewise continuous, although they need not be. However, we expect, but do not prove, that the solution of an initial value problem with piecewise continuous cross sections and a piecewise continuous initial condition will in fact be piecewise continuous.

II. ANALYSIS

We wish to consider a bounded convex physical region consisting of at most a finite number of different material subregions, each having a continuous boundary.

Therefore, we require each subregion D_i, $1 \leqslant i \leqslant I$, to be a bounded open set of points in Euclidean space such that for any fixed $\mathbf{x}_i \in D_i$, $|\mathbf{x} - \mathbf{x}_i|$ is a continuous function of $\mathbf{x}$ for $\mathbf{x} \in \partial D_i$ (the boundary of D_i). We set

$$D = \bigcup_{i=1}^{I} D_i$$

and we require $\mathbf{D}$, the closure of D, to be convex.

As a means of introducing the concepts relevant to our study we consider, for the moment, an initial-boundary value problem. Then the time-dependent multigroup neutron transport equation has the form

$$\frac{\partial}{\partial t}\psi(\mathbf{x},\mathbf{\Omega},t) = A(\psi,\mathbf{\Omega},t), \tag{1}$$

where $\psi(\mathbf{x},\mathbf{\Omega},t)$ is a $G \times 1$ vector whose $\mathbf{g}$th component is the neutron density for the $\mathbf{g}$th group, and the multigroup transport operator A is

$$A = T + K, \tag{2a}$$

$$T\mathbf{f}(\mathbf{x},\mathbf{\Omega}) = -(\mathbf{\Omega}\cdot\nabla)\mathsf{V}\cdot\mathbf{f}(\mathbf{x},\mathbf{\Omega}) - \mathsf{V}\cdot\mathbf{\Sigma}_t(\mathbf{x})\cdot\mathbf{f}(\mathbf{x},\mathbf{\Omega}), \tag{2b}$$

$$K\mathbf{f}(\mathbf{x},\mathbf{\Omega}) = \int \mathsf{V}\cdot\mathbf{\Sigma}(\mathbf{x},\mathbf{\Omega}\cdot\mathbf{\Omega}')\cdot\mathbf{f}(\mathbf{x},\mathbf{\Omega}')d^2\Omega'. \tag{2c}$$

In these equations, $\mathbf{x} \in D$ and $\mathbf{\Omega} \in S$ (the surface of the unit sphere). V is a $G \times G$ diagonal velocity matrix, whose $\mathbf{g}$th diagonal component is v_g, the speed of neutrons in the $\mathbf{g}$th energy group. We assume $v_g > 0$, $1 \leqslant g \leqslant G$. $\mathbf{\Sigma}_t(\mathbf{x})$ is a $G \times G$ diagonal matrix whose $\mathbf{g}$th diagonal component is $\sigma_{tg}(\mathbf{x})$, the total cross section for the $\mathbf{g}$th group at the point $\mathbf{x}$. We assume that each $\sigma_{tg}(\mathbf{x})$ is a nonnegative uniformly continuous function of $\mathbf{x}$ in each subregion D_i, $1 \leqslant i \leqslant I$. $\mathbf{\Sigma}(\mathbf{x},\mu)$ is a $G \times G$ matrix, each component of which, $\sigma_{gg'}(\mathbf{x},\mu)$, is a nonnegative uniformly continuous function of $\mathbf{x}$ and μ for each subregion and $\mu \in [-1,1]$. We impose the following physically reasonable requirement: If, at any point $\mathbf{x}_0$, there exists a μ_0 such that $\sigma_{gg'}(\mathbf{x}_0,\mu_0) > 0$, then $\sigma_{gg'}(\mathbf{x}_0,\mu) > 0$ for all $\mu \in [-1,1]$.

To the transport equations (1), (2), we adjoin initial boundary conditions:

$$\mathbf{\Psi}(\mathbf{x},\mathbf{\Omega},0) = \mathbf{h}(\mathbf{x},\mathbf{\Omega}) \quad (\mathbf{x},\mathbf{\Omega}) \in D \times S, \tag{3a}$$

$$\mathbf{\Psi}(\mathbf{x},\mathbf{\Omega},t) = 0, \quad \mathbf{x} \in \partial D, \quad \mathbf{\Omega}\cdot\mathbf{n}(\mathbf{x}) < 0. \tag{3b}$$

Here $\mathbf{n}(\mathbf{x})$ is the unit outer normal at the point $\mathbf{x} \in \partial D$.

We require the initial condition $\mathbf{h}$ to be an element of H, the Hilbert space of square integrable vector functions defined for $(\mathbf{x},\mathbf{\Omega}) \in D \times S$, with the usual L_2 norm

$$||\mathbf{h}|| = \left[\sum_{g=1}^{G}\int_D\int_S |h_g(\mathbf{x},\mathbf{\Omega})|^2 d^2\Omega d^3x\right]^{1/2}.$$

We seek a solution ψ of (1)–(3) such that for each $t \geqslant 0$, $\psi \in H$, i.e.,

$$||\psi||(t) < \infty, \quad 0 \leqslant t < \infty.$$

This solution exists and is given uniquely by[2]

$$\psi(\mathbf{x},\mathbf{\Omega},t) = E(t)\mathbf{h}(\mathbf{x},\mathbf{\Omega}),$$

where $E(t)$: $H \to H$ is the strongly continuous semigroup whose infinitesimal generator is A, the transport operator defined on a suitable dense subset of H. [One often sees the descriptive notation $E(t) = e^{tA}$.]

Let us define

$$\alpha = \inf_{\substack{1 \leqslant g \leqslant G \\ \mathbf{x} \in D}} \sigma_{tg}(\mathbf{x}),$$

τ = (maximum diameter of D)/(min v_g)
= maximum escape time for an uncollided neutron,

and

C = the cone of vector functions in H, each component of which is nonnegative (except on a set of measure zero).

We say that an operator is *positive* if it maps C into C.

Then the following two results are due to Jörgens[2]:

Theorem 1: $E(t)$: $H \to H$ is a positive operator for $t \geqslant 0$ and is a completely continuous operator for $t > 3\tau$.

Theorem 2: The spectrum of A, $\sigma[A]$, is discrete, situated in the half-plane $\mathrm{Re}\lambda \leqslant ||K|| - \alpha$, and finite in any strip $\beta_1 \leqslant \mathrm{Re}\lambda \leqslant \beta_2$. The generalized eigenspaces are finite-dimensional. Let $\lambda_0, \lambda_1, \lambda_2, \cdots$ be the enumeration of $\sigma[A]$. Then $\sigma[E(t)]$ consists of the points $e^{\lambda_j t}$ plus possibly the point zero for $0 < t < \infty$. The generalized eigenspace of $E(t)$ corresponding to the eigenvalue z is the direct sum of the generalized eigenspaces of A corresponding to all eigenvalues λ such that $e^{\lambda t} = z$.

Jörgens proves Theorem 1 in a manner which we shall not discuss here, and then he proves Theorem 2 by using the results that for $t > 3\tau$, $\sigma[E(t)] - \{0\}$ is discrete, the generalized eigenspaces are finite-dimensional, $\sigma[E(t)]$ has at most an accumulation point at $\lambda = 0$, and

$$\sigma[E(t)] - \{0\} = \{e^{\lambda t} | \lambda \in \sigma[A]\}.$$

Theorem 2 provides a general qualitative description of $\sigma[A]$. In the remainder of this article we go beyond this result to consider the existence of eigenvalues in $\sigma[A]$ and of leading or dominant eigenvalues, as defined in the Introduction. In our analysis we shall make considerable use of the semigroup $E(t)$ and its properties. The idea for this is due to Angelescu and Protopopescu[7] and, of course, Jörgens.[2]

First we consider the existence of points in $\sigma[A]$. The following theorem states that a sufficient condition for this is that there be self-scattering in at least one energy group.

Theorem 3: For at least one energy group g_i, let there

exist an open set $R \subset D$ such that $\sigma_{gg_i}(x,\mu) > 0$ for $\mathbf{x} \in R$ and $\mu \in [-1,1]$. Then $\sigma[A]$ contains at least one point.

Proof: Let R_0 be a closed sphere contained in R,

$$\sigma_t = \max_{\mathbf{x} \in D} \sigma_{tg_i}(\mathbf{x})$$

and

$$\sigma_s = 4\pi \min_{\substack{\mathbf{x} \in R_0 \\ -1 \leqslant \mu \leqslant 1}} \sigma_{gg_i}(x,\mu).$$

Then the g_i th component $\psi_{g_i}(\mathbf{x},\mathbf{\Omega})$ of the solution of the initial value problem (1)–(3) is greater than or equal to the solution $\psi(\mathbf{x},\mathbf{\Omega},t)$ of the problem:

$$\frac{\partial}{\partial t}\psi = -v_t \mathbf{\Omega} \cdot \nabla \psi - v_t \sigma_t \psi + \frac{v_t \sigma_s}{4\pi} \chi_0(\mathbf{x}) \int \psi d^2\Omega', \quad (4a)$$

$$\psi(\mathbf{x},\mathbf{\Omega},0) = h_{g_i}(\mathbf{x},\mathbf{\Omega})\chi_0(\mathbf{x}), \quad \mathbf{x} \in D, \quad (4b)$$

$$\psi(\mathbf{x},\mathbf{\Omega},t) = 0, \quad \mathbf{x} \in \partial D, \quad \mathbf{\Omega} \cdot \mathbf{n}(\mathbf{x}) < 0. \quad (4c)$$

Here

$$\chi_0(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in R_0, \\ 0, & \mathbf{x} \notin R_0, \end{cases}$$

is the characteristic function for R_0.

Physically, problem (4) corresponds to a one-group, homogeneous self-scattering sphere surrounded by a pure absorber. In the Appendix, we show that if the initial condition is positive on a set of finite measure in $R_0 \times S$, then there exist constants a and b, $a > 0$, such that

$$\inf_{\mathbf{x} \in D} \int \psi(\mathbf{x},\mathbf{\Omega},t) d^2\Omega \geqslant ae^{bt}, \quad t \geqslant \tau.$$

Since ψ is (pointwise) less than $\psi_{g_i}(\mathbf{x},\mathbf{\Omega},t)$, it follows that

$$||\psi||(t) \geqslant ke^{bt}, \quad (5)$$

where k is a positive constant. This shows that $\sigma[A]$ contains a finite point; for, otherwise, ψ would decay faster than any exponential function, and this would contradict Eq. (5). Q.E.D.

Next we show that there can exist problems for which $\sigma[A]$ contains no points.

Theorem 4: If $\mathbf{\Sigma}(\mathbf{x},\mu) = 0$, then $\sigma[A]$ is empty.

Proof: $\mathbf{\Sigma} = 0$ corresponds to a purely absorbing medium. For any initial value problem (1)–(3), $\psi(\mathbf{x},\mathbf{\Omega},t) = \mathbf{0}$ for $t > \tau$, and so no solution of the form

$$\psi(\mathbf{x},\mathbf{\Omega},t) = e^{\lambda t}\psi_\lambda(\mathbf{x},\mathbf{\Omega})$$

can exist. Hence there can exist no points in $\sigma[A]$. Q.E.D.

By using a similar argument, one can also show that $\sigma[A]$ is empty if all the components of $\mathbf{\Sigma}$ on and above the main diagonal are zero.

In Theorem 3 we gave a simple condition for the existence of points in $\sigma[A]$. In the next theorem we show that this condition suffices for the existence of a leading eigenvalue.

Theorem 5: If $\sigma[A]$ is nonempty, then it contains a leading eigenvalue.

Proof: Since $\sigma[A]$ contains a finite eigenvalue λ, then $E(t)$ contains the nonzero eigenvalue $e^{\lambda t}$ for all $t > 0$. Thus, for $t > 3\tau$, $E(t)$ is positive, completely continuous, and has a point in its spectrum different from zero. A result of Krein and Rutman[8] now states that $E(t)$ has a positive eigenvalue $\rho(t)$, not less in magnitude than any other characteristic vector, and to $\rho(t)$ corresponds at least one eigenvector $\psi \in C$.

It follows that ψ is an eigenvector of A corresponding to an eigenvalue λ_0, which satisfies

$$e^{\lambda_0 t} = \rho(t).$$

Since $\rho(t) > 0$ for all t, λ_0 must be real. Also, since $\rho(t)$ is not less than the magnitude of any other eigenvalue of $E(t)$, it follows that

$$|e^{\lambda t}| \leqslant \rho(t) = e^{\lambda_0 t}, \quad \lambda \in \sigma[A],$$

and therefore $\mathrm{Re}\lambda \leqslant \lambda_0$ for all $\lambda \in \sigma[A]$. Q.E.D.

The next theorem states, in physical terms, that if the hypothesis of Theorem 5 holds, then a sufficient condition for the leading eigenvalue λ_0 to be dominant is that any neutron which is introduced into the system at any spatial point and in any energy group must have a positive probability of eventually being scattered into any other energy group. This condition is not necessary, as we show in Sec. III.

Theorem 6: Let the hypothesis of Theorem 3 hold. Then $\sigma[A]$ possesses a leading eigenvalue λ_0. This eigenvalue is dominant if the following condition holds: For each point $\mathbf{x}^* \in D$ and any two energy groups g_1^*, g_2^*, there exists a sequence of integers

$$g_1, g_2, \ldots, g_n, \quad 1 \leqslant g_i \leqslant G,$$

$$g_1 = g_1^*, \quad g_n = g_2^*,$$

such that

$$\sigma_{g_1 g_1}(\mathbf{x}^*,\mu) > 0$$

and

$$\max_{\substack{\mathbf{x} \in D \\ -1 \leqslant \mu \leqslant 1}} \sigma_{g_i g_{i-1}}(x,\mu) > 0, \quad 2 \leqslant i \leqslant n-1.$$

Moreover, if these conditions hold, then the eigenfunction $\psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega})$ is the only nonnegative eigenfunction of A.

Proof: Let $\mathbf{h}(\mathbf{x},\mathbf{\Omega}) \in C$, and let us consider the initial value problem (1)–(3) with initial condition $\mathbf{h}$.

Since the hypothesis of Theorem 3 holds, there exists a g_i and a closed sphere $R_0 \subset D$ for which $\sigma_{gg_i}(x,\mu) > 0$, $\mathbf{x} \in R_0$.

By the assumed conditions on $\mathbf{\Sigma}$, it is clear that there exists a time t_0, $0 \leqslant t_0 \leqslant G\tau$, for which $\psi_{g_i}(\mathbf{x},\mathbf{\Omega},t_0)$ is positive on a set of positive measure in $R_0 \times S$. The results of the Appendix imply that

$$\inf_{\mathbf{x} \in D} \int \psi_{g_i}(\mathbf{x},\mathbf{\Omega},t) d^2\Omega > 0, \quad t > (G+1)\tau.$$

The $\mathbf{g}_i$ th group neutrons thus act as a positive neutron source which, by the assumptions on $\mathbf{\Sigma}$, eventually feed all other groups. This leads to the result

$$\inf_{\substack{\mathbf{x} \in D \\ 1 \leqslant g \leqslant G}} \int \psi_g(\mathbf{x},\mathbf{\Omega},t) d^2\Omega > m(t) > 0, \quad t > 2G\tau. \quad (6)$$

It can also be shown that, for $t > 2\tau$, $\psi(\mathbf{x},\mathbf{\Omega},t)$ is bounded. [In fact, if the initial condition $\mathbf{h}(\mathbf{x},\mathbf{\Omega})$ is bounded, then ψ is bounded for $t \geqslant 0$.] The method[9] is to write $\psi = \Sigma_{n=0}^{\infty}\psi_n$, where ψ_n are the nth collided neutrons, and to show that ψ_n is bounded for $n \geqslant 2$. Then, for $t > 2\tau$, $\psi_0 = \psi_1 = 0$, and so ψ is bounded. We shall not give the details here. However, this

result implies

$$\sup_{\substack{\mathbf{x}\in D\\ 1\leqslant g\leqslant G}} \int \psi_g(\mathbf{x},\mathbf{\Omega},t)d^2\Omega < M(t) < \infty, \quad t > 2\tau. \tag{7}$$

Equations (2c), (6), and (7) now give, for $t > 2G\tau$,

$$v_g m(t)\left[\sum_{j=1}^{G} \inf_{\mu}\sigma_{gj}(\mathbf{x},\mu)\right] \leqslant K\psi_g(\mathbf{x},\mathbf{\Omega},t) \leqslant v_g M(t)\left[\sum_{j=1}^{G} \sup_{\mu}\sigma_{gj}(\mathbf{x},\mu)\right].$$

(Here $K\psi_g$ is the gth component of $K\psi$, and $1\leqslant g\leqslant G$.) By the assumptions on $\mathbf{\Sigma}$ originally stated in this section, $\inf_\mu \sigma_{gj}(\mathbf{x},\mu)$ and $\sup_\mu \sigma_{gj}(\mathbf{x},\mu)$ are either both zero or both nonzero, for all g, j, and $\mathbf{x}$. Thus there exists a nonnegative function $\chi_g(\mathbf{x})$ which is continuous in each subregion D_i and positive constants α_g, β_g such that

$$\frac{1}{\beta_g}\left[\sum_{j=1}^{G} \sup_{\mu}\sigma_{gj}(\mathbf{x},\mu)\right] \leqslant \chi_g(\mathbf{x}) \leqslant \frac{1}{\alpha_g}\left[\sum_{j=1}^{G} \inf_{\mu}\sigma_{gj}(\mathbf{x},\mu)\right],$$

and hence

$$\alpha_g v_g m(t)\chi_g(\mathbf{x}) \leqslant K\psi_g(\mathbf{x},\mathbf{\Omega},t) \leqslant \beta_g v_g M(t)\chi_g(\mathbf{x}), \quad 1\leqslant g\leqslant G, \tag{8}$$

for all $(\mathbf{x},\mathbf{\Omega})\in D\times S$ and $t > 2G\tau$.

Now let

$$t_0 = 2G\tau, \quad t_1 = t_0 + \tau,$$

$$m = \inf_{t_0\leqslant t\leqslant t_1} m(t), \quad M = \sup_{t_0\leqslant t\leqslant t_1} M(t).$$

Then, for $t_0\leqslant t\leqslant t_1$, $\psi_g(\mathbf{x},\psi,t)$ is bounded from above and below by

$$\xi_g^{(0)}(\mathbf{x},\mathbf{\Omega},t) \leqslant \psi_g(\mathbf{x},\mathbf{\Omega},t) \leqslant \xi_g^{(1)}(\mathbf{x},\mathbf{\Omega},t), \tag{9}$$

where $\xi_g^{(n)}$, $n = 0, 1$, satisfies

$$\frac{\partial}{\partial t}\xi_g^{(n)} + v_g\mathbf{\Omega}\cdot\nabla\xi_g^{(n)} + v_g\sigma_{tg}(\mathbf{x})\xi_g^{(n)} = \gamma^{(n)}\chi_g(\mathbf{x}), \tag{10a}$$

$$\xi_g^{(n)}(\mathbf{x},\mathbf{\Omega},t_0) = \psi_g(\mathbf{x},\mathbf{\Omega},t_0), \quad \mathbf{x}\in D, \tag{10b}$$

$$\xi_g^{(n)}(\mathbf{x},\mathbf{\Omega},t) = 0, \quad t > t_0, \ \mathbf{x}\in\partial D, \ \mathbf{\Omega}\cdot\mathbf{n} < 0, \tag{10c}$$

and

$$\gamma^{(0)} = \alpha_g v_g m, \tag{11a}$$

$$\gamma^{(1)} = \beta_g v_g M. \tag{11b}$$

Solving problem (10) at $t = t_1$ gives

$$\xi_g^{(n)}(\mathbf{x},\mathbf{\Omega}) = \gamma^{(n)}\xi_g(\mathbf{x},\mathbf{\Omega}),$$

where

$$\xi_g(\mathbf{x},\mathbf{\Omega}) = \int_0^{d(\mathbf{x},\mathbf{\Omega})} \chi_g(\mathbf{x}-s\mathbf{\Omega}) \times\exp\left[-\int_0^s v_g\sigma_{tg}(\mathbf{x}-s'\mathbf{\Omega})ds'\right]ds$$

and where $d(\mathbf{x},\mathbf{\Omega})$ is the distance from $\mathbf{x}$ to ∂D in the direction of $-\mathbf{\Omega}$. Equation (9) becomes at $t = t_1$,

$$(\alpha_g v_g m)\xi_g(\mathbf{x},\mathbf{\Omega}) \leqslant \psi_g(\mathbf{x},\mathbf{\Omega},t_1) \leqslant (\beta_g v_g M)\xi_g(\mathbf{x},\mathbf{\Omega}).$$

Since $\alpha_g v_g m > 0$ and $\xi_g \geqslant 0$, we have shown the following: There exists a vector function $\xi\in C$ such that, for any nonzero function $\mathbf{h}\in C$, there exist positive constants α and β such that

$$\alpha\xi \leqslant E(t_1)\mathbf{h} \leqslant \beta\xi, \tag{12}$$

where the operator $E(t_1)$ is completely continuous. [The inequalities (12), of course, hold componentwise.] Three results of Krasnosel'skii[10] now imply that $e^{\lambda_0 t_1}$ is a simple eigenvalue of $E(t_1)$ which is greater than the magnitude of any other eigenvalue and that the corresponding eigenfunction ψ_{λ_0} is the only eigenfunction of $E(t_1)$ in C.

It follows directly that λ_0 is a real simple eigenvalue of A which is strictly greater than the real part of any other eigenvalue of A. The corresponding eigenfunction ψ_{λ_0} is the only eigenfunction of A which lies in the cone C, and the initial value problem with initial condition $\mathbf{h} = \psi_{\lambda_0}$ has the solution

$$\psi(\mathbf{x},\mathbf{\Omega},t) = e^{\lambda_0 t}\psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega}),$$

which satisfies Eq. (6). Therefore,

$$\int \psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega})d^2\Omega$$

is a vector function of $\mathbf{x}$ which is everywhere positive in D. This completes the proof of the theorem. Q.E.D.

To illustrate this theorem, we shall consider some simple generic examples. For a homogeneous, two-group medium,

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

does not satisfy the conditions of the theorem because group 2 neutrons cannot scatter into group 1. (In this and all the following examples, the 1's can be replaced by arbitrary positive numbers.) However,

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

does satisfy the conditions of the theorem, as does any n-group $\mathbf{\Sigma}$ with all positive components. Also, in G groups,

$$\mathbf{\Sigma} = \begin{pmatrix} 0 & \cdots & & & 1 \\ 1 & 0 & & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & & 0 & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

satisfies the conditions of the theorem because after $G-1$ collisions a neutron in any group will scatter through all others.

Now let us consider a two-group region D consisting of two material subregions D_1, D_2, for which

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

In D_1, group 2 neutrons cannot scatter into group 1. However, any group 2 neutron in D_1 can be scattered into D_2, where it can undergo a collision and then be scattered into group 1. Therefore, the conditions of Theorem 6 are met. They are also met for

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

but not for

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

since a group 1 neutron in D_1 will stream out of D or be absorbed in D_1 (but in either case will not be scattered) if it travels along a path which does not intersect D_2.

The following (and final) theorem provides some simple relationships between the leading eigenvalue of the multigroup operator A, defined in Eqs. (2), and the leading eigenvalues of the G one-group operators A_g, $1 \leqslant g \leqslant G$, defined by

$$A_g f(\mathbf{x},\mathbf{\Omega}) = -\mathbf{\Omega}\cdot\nabla v_g f(\mathbf{x},\mathbf{\Omega}) - v_g \sigma_{tg}(\mathbf{x}) f(\mathbf{x},\mathbf{\Omega}) + v_g \int \sigma_{gg}(\mathbf{x},\mathbf{\Omega}\cdot\mathbf{\Omega}') f(\mathbf{x},\mathbf{\Omega}') d^2\Omega'.$$

[These operators are obtained by setting to zero the off-diagonal components of $\mathbf{\Sigma}$, yielding a G-group diagonal transport operator A' whose gth diagonal component is the operator A_g.]

Theorem 7: Let λ_0 be the leading eigenvalue of A and λ_{0g} be the leading eigenvalue of A_g. (If any one of these operators has no finite spectrum, we define the dominant eigenvalue to be $-\infty$.) Then

$$\max_{1 \leqslant g \leqslant G} \lambda_{0g} \leqslant \lambda_0. \tag{13}$$

Moreover, if $\mathbf{\Sigma}$ is a lower triangular matrix, i.e., $\sigma_{ij} = 0$ for $i < j$, then

$$\max_{1 \leqslant g \leqslant G} \lambda_{0g} = \lambda_0. \tag{14}$$

Proof: Let $\mathbf{\Sigma}'$ be the matrix whose diagonal components are equal to those of $\mathbf{\Sigma}$ and whose off-diagonal components are zero. Let A' be the transport operator A with $\mathbf{\Sigma}$ replaced by $\mathbf{\Sigma}'$. Then the solution ψ of the initial value problem (1)–(3) is greater than or equal to ψ', the solution of the same problem but with $\mathbf{\Sigma}$ replaced by $\mathbf{\Sigma}'$, provided the initial condition $\mathbf{h} \in C$.

Let $\lambda_{0k} = \max_{1 \leqslant g \leqslant G} \lambda_{0g}$, and let $\psi_{\lambda_{0k}}(\mathbf{x},\mathbf{\Omega})$ be the corresponding eigenfunction. Then the vector $\mathbf{h}$ with $\psi_{\lambda_{0k}}$ as its kth component and all other components zero is an element of the cone C, and the solutions ψ and ψ' of the stated initial value problems with this particular $\mathbf{h}$ satisfy

$$\psi_k(\mathbf{x},\mathbf{\Omega},t) \geqslant \psi'_k(\mathbf{x},\mathbf{\Omega},t) = e^{\lambda_{0k}t} \psi_{\lambda_{0k}}(\mathbf{x},\mathbf{\Omega}).$$

Therefore, ψ_k (and hence ψ) decays no faster than $e^{\lambda_{0k}t}$, and so $\lambda_0 \geqslant \lambda_{0k}$.

Next, we let $\mathbf{\Sigma}$ be lower triangular. If $\lambda_0 = -\infty$, then, by Eq. (13), $\lambda_{0g} = -\infty$, $1 \leqslant g \leqslant G$, and so Eq. (14) holds. If λ_0 is finite, let ψ_{λ_0} be the corresponding eigenfunction in C, and let $\psi_{\lambda_0 k}$ be its first nonzero component. Then $\psi_{\lambda_0 k} \geqslant 0$, and since $\mathbf{\Sigma}$ is lower triangular, λ_0 is the dominant eigenvalue (and $\psi_{\lambda_0 k}$ the corresponding eigenfunction) of the operator A_k. Thus, $\lambda_0 = \lambda_{0k}$, and Eq. (14) then follows from Eq. (13). Q.E.D.

III. EXAMPLES

Here we present some simple examples to show that if the conditions of Theorem 6 are not met, then a dominant eigenvalue λ_0 need not exist, i.e., λ_0 need not be simple and its eigenfunction(s) need not be positive. On the other hand, we also show that if these conditions are not met, then a dominant eigenvalue λ_0 *can* exist.

Let D be a finite convex domain, and let λ_0 be the dominant eigenvalue and ψ_{λ_0} the corresponding eigenfunction of the one-group transport operator

$$A_1\psi(\mathbf{x},\mathbf{\Omega}) = -\mathbf{\Omega}\cdot\nabla\psi(\mathbf{x},\mathbf{\Omega}) - \psi(\mathbf{x},\mathbf{\Omega}) + \frac{c}{4\pi} \times \int \psi(\mathbf{x},\mathbf{\Omega}') d^2\Omega'$$

with vacuum boundary conditions. Let A be the two-group operator

$$A \begin{bmatrix} \psi_1(\mathbf{x},\mathbf{\Omega}) \\ \psi_2(\mathbf{x},\mathbf{\Omega}) \end{bmatrix} = -(\mathbf{\Omega}\cdot\nabla) \begin{bmatrix} \psi_1(\mathbf{x},\mathbf{\Omega}) \\ \psi_2(\mathbf{x},\mathbf{\Omega}) \end{bmatrix} - \begin{bmatrix} \psi_1(\mathbf{x},\mathbf{\Omega}) \\ \psi_2(\mathbf{x},\mathbf{\Omega}) \end{bmatrix} + \frac{1}{4\pi} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \int \begin{bmatrix} \psi_1(\mathbf{x},\mathbf{\Omega}') \\ \psi_2(\mathbf{x},\mathbf{\Omega}') \end{bmatrix} d^2\Omega',$$

defined in D, also with vacuum boundary conditions.

First let us take

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}.$$

Then A corresponds to a slowing down medium, for which the scattering matrix $\mathbf{\Sigma}$ is commonly lower triangular. Here (and for all lower triangular $\mathbf{\Sigma}$) the conditions of Theorem 6 are not met because group 2 neutrons cannot scatter into group 1.

If $a \leqslant c$, then the dominant eigenvalue of A is λ_0, with the nonpositive eigenfunction

$$\mathbf{\psi}_{\lambda_0} = \begin{pmatrix} 0 \\ \psi_{\lambda_0} \end{pmatrix}.$$

If $a = c$, then the dominant eigenvalue of A is still λ_0, but the generalized eigenspace is now two-dimensional. This space is spanned by

$$\mathbf{\psi}_{\lambda_0} = \begin{pmatrix} 0 \\ \psi_{\lambda_0} \end{pmatrix} \quad \text{and} \quad \mathbf{\psi}^{(1)}_{\lambda_0} = \begin{pmatrix} \psi_{\lambda_0} \\ bf \end{pmatrix},$$

where $f(\mathbf{x},\mathbf{\Omega})$ is any solution of the problem

$$(\lambda_0 I - A_1) f = -(I - P) \frac{1}{4\pi} \int \psi_{\lambda_0} d^2\Omega'$$

and where P is the projection operator

$$Pg(\mathbf{x},\mathbf{\Omega}) = \psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega}) \frac{\int \psi^*_{\lambda_0}(\mathbf{x}',\mathbf{\Omega}') g(\mathbf{x}',\mathbf{\Omega}') d^2\Omega' d^3x'}{\int \psi^*_{\lambda_0}(\mathbf{x}',\mathbf{\Omega}') \psi_{\lambda_0}(\mathbf{x}',\mathbf{\Omega}') d^2\Omega' d^3x'}.$$

For $b > 0$, only $\mathbf{\psi}_{\lambda_0}$ is an eigenfunction; $\mathbf{\psi}^{(1)}_{\lambda_0}$ is a generalized eigenfunction satisfying $(\lambda_0 I - A)\mathbf{\psi}^{(1)}_{\lambda_0} = b a_0 \mathbf{\psi}_{\lambda_0}$, where a_0 is a constant. It is clear that $\mathbf{\psi}_{\lambda_0}$ is nonpositive, but little can be said about the positivity of $\mathbf{\psi}^{(1)}_{\lambda_0}$.

Now let us take

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} c & 0 \\ a & b \end{pmatrix}.$$

with $0 < a$ and $0 \leqslant b < c$. Again, the conditions of Theorem 6 are not met, but now A does possess the dominant eigenvalue

λ_0, with eigenfunction

$$\psi_{\lambda_0} = \begin{pmatrix} \psi_{\lambda_0} \\ \xi \end{pmatrix},$$

where ξ satisfies the equation

$$(\lambda_0 - A_2)\xi(\mathbf{x},\mathbf{\Omega}) = \mathbf{\Omega}\cdot\nabla\xi + (1+\lambda_0)\xi - \frac{b}{4\pi}\int \xi d^2\Omega'$$
$$= \frac{a}{4\pi}\int \psi_{\lambda_0} d^2\Omega'$$

with vacuum boundary conditions. Since $b < c$, the operator acting on ξ corresponds to a subcritical medium, and since $\int \psi_\lambda d^2\Omega' > 0$, the "source" driving ξ is positive. Hence ξ exists and is positive, except, of course, on the boundary for incident directions, and thus this is also true for ψ_{λ_0}.

APPENDIX

In this appendix we prove two lemmas pertaining to the analysis in Sec. II. Lemma 1 is required for the proof of Lemma 2, and Lemma 2 is used directly in the proofs of Theorems 5 and 6 in Sec. II. The lemmas are:

Lemma 1: The one-group transport operator A for a homogeneous, isotropically scattering sphere R_0 has a real eigenvalue λ_0, to which corresponds an eigenfunction $\psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega})$ satisfying

(a) ψ_{λ_0} is continuous and bounded in $\bar{R}_0 \times S$,

(b) $\inf_{\mathbf{x}\in R_0} \int \psi_{\lambda_0}(\mathbf{x}.\mathbf{\Omega}) d^2\Omega > 0$,

(c) $\psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega}) > 0$ for all $\mathbf{x},\mathbf{\Omega}$ such that $(\mathbf{x} - \epsilon\mathbf{\Omega})\in R_0$ for some $\epsilon > 0$.

Lemma 2: Let D be the convex domain described in Sec. II, and let $\sigma_t(\mathbf{x}),\sigma_s(\mathbf{x})$ be nonnegative, piecewise continuous functions of $\mathbf{x}$ with $\sigma_s(\mathbf{x}) > 0$ on a set $D_0 \subset D$ of positive measure. Let ψ be the solution of the problem

$$\frac{\partial}{\partial t}\psi = A\psi,$$

$$Af(\mathbf{x},\mathbf{\Omega}) = -v\mathbf{\Omega}\cdot\nabla f(\mathbf{x},\mathbf{\Omega}) - v\sigma_t(\mathbf{x}) f(\mathbf{x},\mathbf{\Omega}) + \frac{v\sigma_s(\mathbf{x})}{4\pi}\int f(\mathbf{x},\mathbf{\Omega}') d^2\Omega',$$

$$\psi(\mathbf{x},\mathbf{\Omega},0) = h(\mathbf{x},\mathbf{\Omega}),$$

$$\psi(\mathbf{x},\mathbf{\Omega},t) = 0, \quad \mathbf{x}\in\partial D, \ \mathbf{\Omega}\cdot\mathbf{n} < 0,$$

where h is nonnegative in $D \times S$ and positive on a set of positive measure in $D_0 \times S$. Then there exists a positive constant a and a real constant b such that

$$\inf_{\mathbf{x}\in D} \int \psi(\mathbf{x},\mathbf{\Omega},t) d^2\Omega > ae^{bt}, \quad t \geqslant 3\tau.$$

Proof of Lemma 1: Van Norton[3] has shown that A possesses an infinite number of real eigenvalues. Thus by Theorem 3 (Sec. 2), A possesses a real eigenvalue λ_0, to which there corresponds a nonnegative eigenfunction ψ_{λ_0}. Since the sphere is homogeneous, ψ_{λ_0} is continuous[2] and hence bounded. The scalar flux

$$\rho_{\lambda_0}(\mathbf{x}) = \int \psi_{\lambda_0}(\mathbf{x},\mathbf{\Omega}) d^2\Omega$$

is nonnegative and satisfies an integral equation with an everywhere positive kernel[3]; hence,

$$\inf_{\mathbf{x}\in R_0} \rho_{\lambda_0}(\mathbf{x}) > 0.$$

Finally, let σ_t and σ_s be the (constant) total and scattering cross sections, and $d(\mathbf{x},\mathbf{\Omega})$ denote the distance from $\mathbf{x}$ to ∂R_0 in the direction of $-\mathbf{\Omega}$. Then

$$\psi(\mathbf{x},\mathbf{\Omega}) = \int_0^{d(\mathbf{x},\mathbf{\Omega})} e^{-(v\sigma_t + \lambda_0)s} \frac{\sigma_s}{4\pi}\rho(\mathbf{x} - s\mathbf{\Omega}) ds,$$

and thus $\psi > 0$ for all $(\mathbf{x},\mathbf{\Omega})$ such that $d(\mathbf{x},\mathbf{\Omega}) > 0$. This completes the proof of the lemma.

Proof of Lemma 2: Functions in $B = L_2(D \times S)$ belong to equivalence classes, i.e., ψ_1 and ψ_2 are in the same equivalence class if $||\psi_1 - \psi_2|| = 0$. This means that $\psi_1(\mathbf{x},\mathbf{\Omega}) = \psi_2(\mathbf{x},\mathbf{\Omega})$ for all $(\mathbf{x},\mathbf{\Omega})\in D \times S$, except for a set of measure zero.

Consider an initial value problem for the transport operator A, with two initial conditions h_1 and h_2 in the same equivalence class. Let the solutions be ψ_1 and ψ_2. Then $\psi_i = E(t)h_i$, $i = 1,2$, and so

$$\psi_1 - \psi_2 = E(t)(h_1 - h_2),$$

whose $h_1 - h_2 = 0$, except for a set of measure zero.

For $t > 3\tau$, $E(t)$ is completely continuous, and hence is an integral operator.[2] Since an integral operator acting on a function which is zero (except for a set of measure zero) is zero, it follows that

$$\psi_1(\mathbf{x},\mathbf{\Omega}) = \psi_2(\mathbf{x},\mathbf{\Omega}), \quad (\mathbf{x},\mathbf{\Omega})\in D \times S, \quad t > 3\tau.$$

Now let us turn to the problem at hand. By the above remarks, we can assume $h > 0$ on an open set in $D_0 \times S$. We write

$$\psi = \psi_0 + \hat{\Psi},$$

where ψ_0 are the uncollided neutrons. Since the source which feeds $\hat{\Psi}$ is isotropic, there exists a sphere $R_0 \subset D_0$ and a (small) time t_0 for which

$$\inf_{\substack{\mathbf{x}\in R_0 \\ |\mathbf{\Omega}| = 1}} \hat{\Psi}(\mathbf{x},\mathbf{\Omega},t_0) \geqslant m_0 > 0. \tag{A1}$$

Now let z_0 and $\tilde{\psi}_{z_0}(\mathbf{x},\mathbf{\Omega})$ be the dominant eigenvalue and eigenfunction, described in Lemma 1, for the following transport operator $\mathscr{A}_0$, defined on the sphere R_0:

$$\mathscr{A}_0 f(\mathbf{x},\mathbf{\Omega}) = -v\mathbf{\Omega}\cdot\nabla f(\mathbf{x},\mathbf{\Omega}) - v\sigma_t f(\mathbf{x},\mathbf{\Omega}) + \frac{v\sigma_s}{4\pi} \times \int f(\mathbf{x},\mathbf{\Omega}') d^2\Omega', \quad \mathbf{x}\in R_0,$$

$$\sigma_t = \sup_{\mathbf{x}\in D}\sigma_t(\mathbf{x}), \quad \sigma_s = \inf_{\mathbf{x}\in R_0}\sigma_s(\mathbf{x}).$$

Since $\tilde{\psi}_{z_0}$ is bounded, there exists a positive constant m_1 such that $m_0 > m_1\tilde{\psi}_{z_0}(\mathbf{x},\mathbf{\Omega})$, $\mathbf{x}\in R_0$. Then by Eq. (A1),

$$\hat{\Psi}(\mathbf{x},\mathbf{\Omega},t_0) \geqslant m_1\tilde{\psi}_{z_0}(\mathbf{x},\mathbf{\Omega}), \quad \mathbf{x}\in R_0.$$

Therefore,

$$\hat{\Psi}(\mathbf{x},\mathbf{\Omega},t) \geqslant \Psi(\mathbf{x},\mathbf{\Omega},t), \quad \mathbf{x}\in D, \ t \geqslant t_0,$$

where Ψ is the solution of the problem

$$\frac{\partial}{\partial t}\Psi = \mathscr{A}\Psi, \quad \mathbf{x}\in D, \ t \geqslant t_0,$$

$$\mathscr{A}f(\mathbf{x},\mathbf{\Omega}) = -v\mathbf{\Omega}\cdot\nabla f(\mathbf{x},\mathbf{\Omega}) - v\sigma_t f(\mathbf{x},\mathbf{\Omega}) + \frac{v\sigma_s \chi_0(\mathbf{x})}{4\pi}\int f(\mathbf{x},\mathbf{\Omega}')d^2\Omega',$$

$$\psi(\mathbf{x},\mathbf{\Omega},t_0) = n_1 \tilde{\psi}_{z_a}(\mathbf{x},\mathbf{\Omega})\chi_0(\mathbf{x}), \quad \mathbf{x}\in D$$

$$\psi(\mathbf{x},\mathbf{\Omega},t) = 0, \quad \mathbf{x}\in\partial D, \ \mathbf{\Omega}\cdot\mathbf{n} < 0, \ t \geqslant t_0,$$

and where $\chi_0(\mathbf{x})$ is the characteristic function for R_0:

$$\chi_0(\mathbf{x}) = \begin{cases} 1, & \mathbf{x}\in R_0, \\ 0, & \mathbf{x}\notin R_0. \end{cases}$$

In this problem, D consists of a homogeneous, self-scattering sphere R_0 surrounded by a pure absorber. For $\mathbf{x}\in R_0$, we have by construction

$$\psi(\mathbf{x},\mathbf{\Omega},t) = m_1\tilde{\psi}_{z_0}(\mathbf{x},\mathbf{\Omega})e^{z_0 t}.$$

For $\mathbf{x}\notin R_0$, ψ can be constructed by tracing characteristics. If $\mathbf{\Omega}$ is such that the line $\mathbf{x} - s\mathbf{\Omega}$, $s > 0$, does not intersect R_0, then $\psi(\mathbf{x},\mathbf{\Omega},t) = 0$. If the line $\mathbf{x} - s\mathbf{\Omega}$, does intersect R_0, and if $d_0(\mathbf{x},\mathbf{\Omega})$ is the minimum value of s for which $\mathbf{x} - s\mathbf{\Omega}$, intersects ∂R_0 (the surface of R_0), then $\psi(\mathbf{x},\mathbf{\Omega},t) = 0$ for $t_0 \leqslant t < t_0 + v^{-1}d_0(\mathbf{x},\mathbf{\Omega})$, while, for $t \geqslant t_0 + v^{-1}d_0(\mathbf{x},\mathbf{\Omega})$, we get

$$\psi(\mathbf{x},\mathbf{\Omega},t) = m_1\left\{\exp\left[-\left(\sigma_t + \frac{z_0}{v}\right)d_0(\mathbf{x},\mathbf{\Omega})\right] \times \tilde{\psi}_{z_a}[\mathbf{x} - d_0(\mathbf{x},\mathbf{\Omega})\mathbf{\Omega},\mathbf{\Omega}]\right\}e^{\lambda_0 t}. \tag{A2}$$

Therefore, for $t > t_0 + \tau$, we can write

$$\psi(\mathbf{x},\mathbf{\Omega},t) = m_1\psi_{z_0}(\mathbf{x}.\mathbf{\Omega})e^{z_0 t}, \tag{A3}$$

where $\psi_{z_0} = \tilde{\psi}_{z_0}$ for $\mathbf{x}\in R_0$ and for $\mathbf{x}\notin R_0$, ψ_{z_0} is either zero or defined by the term in braces in Eq. (A2), depending on whether the line $\mathbf{x} - s\mathbf{\Omega}$, $s > 0$, intersects R_0.

For any $\mathbf{x}\in D$, there exists a nonempty sector of S for which $\psi_{z_0}(\mathbf{x}.\mathbf{\Omega}) > 0$. Therefore,

$$\inf_{\mathbf{x}\in D}\int \psi_{z_0}(\mathbf{x},\mathbf{\Omega})d^2\Omega = m_2 > 0,$$

and so Eq. (A3) gives

$$\inf_{\mathbf{x}\in D}\int \Psi(\mathbf{x},\mathbf{\Omega},t)d^2\Omega \geqslant m_1 m_2 e^{z_0 t}, \quad t > 2\tau,$$

and since $\psi = \psi_0 + \hat{\Psi} \geqslant \Psi$, it follows that

$$\inf_{\mathbf{x}\in D}\int \psi(\mathbf{x},\mathbf{\Omega},t)d^2\Omega \geqslant m_1 m_2 e^{z_0 t}, \quad t > 2\tau,$$

This completes the proof of the lemma. Q.E.D.

ACKNOWLEDGMENT

The author wishes to thank Dr. Thomas Hill for some helpful discussions.

Note added in proof: Two items relevant to this paper have recently come to the author's attention. First, at the Sixth Conference on Transport Theory (Tucson, April 1979), Professor Jürgen Voigt presented some new results which prove that a leading eigenvalue of the multigroup transport operator is necessarily *greater* than the real part of any other eigenvalue. These results are planned to be included in a future journal article. Second, Y. Ming-zhu and Z. Kuang-tien, in "The Spectrum of Transport Operator with Continuous Energy in Inhomogeneous Medium with any Cavity," Scientia Sinica **21**, 298 (1978), have proved similar results as in Ref. 7. However these authors do not use a semigroup approach, and they consider a region which possibly contains voids.

[1] J. Lehner and G.M. Wing, Comm. Pure Appl. Math. **8**, 217 (1955); Duke Math J. **23**, 125 (1956).
[2] K. Jörgens, Comm. Pure Appl. Math. **11**, 219 (1958).
[3] R. Van Norton, Comm. Pure Appl. Math. **15**, 149 (1962).
[4] S. Albertoni and B. Montagnini, *Proceedings of the Symposium on Pulsed Neutron Research* (IAEA, Vienna, 1965), Vol. 1, p. 239; J. Math. Anal. Appl. **13**, 19 (1966).
[5] I. Vidav, J. Math. Anal. Appl. **22**, 144 (1968).
[6] E.W. Larsen and P.F. Zweifel, J. Math. Phys. **15**, 1987 (1974).
[7] N. Angelescu and V. Protopopescu, Rev. Roum. Phys. **22**, 1055 (1977).
[8] M.G. Krein and M.A. Rutman, *Linear Operators Leaving Invariant a Cone in a Banach Space*, Amer. Math. Soc. Transl., Ser. 1, Vol. 10 (Amer. Math. Soc., Providence, R.I., 1962). The relevant result here is Theorem 6.1.
[9] K.M. Case and P.F. Zweifel, J. Math. Phys. **4**, 1376 (1963).
[10] M.A. Krasnosel'skii, *Positive Solutions of Operator Equations* (Noordhoff, Groningen, The Netherlands, 1964). The relevant results are Theorems 2.10, 2.11, and 2.13.

On the problem of existence of quantum field theory

M. Chaichian
Research Institute for Theoretical Physics, University of Helsinki, Finland

M. Hayashi
Department of Physics, Saitama College, Moroyama, Saitama, Japan

N. F. Nelipa and A. E. Pukhov
Institute of Nuclear Physics, Moscow State University, USSR

(Received 11 December 1978)

Existence of quantum field theory is considered for the four-dimensional φ^3 model. The mathematical tool of contraction mapping principle is used to investigate the question of existence of solution for the infinite system of coupled equations for the Green functions of the theory in the Euclidean region. Formulation of the problem for this model with one divergent part is interesting in itself and provides the first attempt towards the study of other renormalizable quantum field theory models with an infinite number of divergent graphs. For sufficiently small values of coupling constant, the theory has a unique solution for the truncated system of equations for the Green functions. However, for the complete, infinite set of equations, the Banach fixed point theorem admits a solution only when the coupling constant tends to zero. Possible reasons for such a result are discussed.

1. INTRODUCTION

A question of fundamental importance in quantum field theory is the existence of exact, nonperturbative solutions for the Green functions. In fact, even with the renormalization procedure and the remarkable experimental success of quantum electrodynamics with the perturbation expansion, still one does not know whether the set of equations describing the theory is self-consistent and whether an exact solution, in principle, exists. These problems, however, are highly nontrivial and, therefore, any partial result in any other field theory models is of great interest. In this work we confine ourselves to this question for the first nontrivial model in quantum field theory with divergence—the four-dimensional φ^3 theory. The characteristic of such a theory is that it is super-renormalizable, i.e., has only one (self-energy) divergent diagram. We shall exploit the powerful mathematical technique of principle of contraction mapping or Banach's fixed point theorem[1] and try to find conditions under which the full, as well as the truncated, system of coupled chain of equations for the vertex functions of four-dimensional $(\lambda/3)\varphi^3$ theory has solution in Euclidean region of variables.[2] In doing so, however, we find that while the truncated system has a unique solution for sufficiently small values of coupling constant λ, the complete set of equations for the Green functions has a solution only if the coupling constant of the theory λ tends to zero. Possible reasons for this result may be the majorization of the integrals and the estimates of the inequalities which appear and/or the definition of norms for different functions. On the other hand, it is known that such field theories as $(\lambda/3)\varphi^3$ have zero radius of convergence in λ, so that it could appear very hard to use the contraction mapping method, however the norms are altered, to make such an iterative solution well-defined. Thus it might be interesting to indicate how other fixed point approaches such as topological degree[3,4] may be of value for such problems. In fact, this method has been preliminarily discussed in the field theory context.[5] A more detailed analysis of these points is apparently necessary and very important to carry out. In particular, some good reason, such as a counterexample, could be looked for in order to rule out the possible expectation that contraction mapping will always fail whenever there is zero radius of convergence of the iterative solution.

The methods of functional analysis have been previously used to discuss the existence of solutions of equations in different branches of theoretical physics.[6-10] In this approach an equation is considered as an operator equation in the Banach space. Such a method has been first suggested and applied to the equations of statistical physics by Bogoliubov *et al.*[6] Subsequently, this approach was applied to different nonlinear equations: to the Chew-Low equations,[7] to the equations for the π scattering off a static nucleon,[8] to the dispersion relations,[9] and to the equations for the Green functions of the φ^3 theory.[10] In Ref. 10 the equation for the two-dimensional φ^3 model were considered, and it was established that Banach's fixed point theorem[1] allows the existence of solutions only for the zero coupling constant. The present work is one step towards a realistic model in four dimensions. Here the occurence of the divergence and its extraction and separation from the quantities which enter the coupled equations is an interesting technical problem and necessary for treating any renormalizable quantum field theory with divergences.

Section 2 contains the derivation of infinite set of coupled nonlinear integral equations for the Green functions from the properties of Feynman diagrams. In Sec. 3 principle of contraction mapping is introduced and exploited for our problem. Section 4 deals with the final results concerning the conditions for existence of solution for Green functions. In Appendix A we give the analytical derivation of

 0022-2488/79/081783-08$01.00

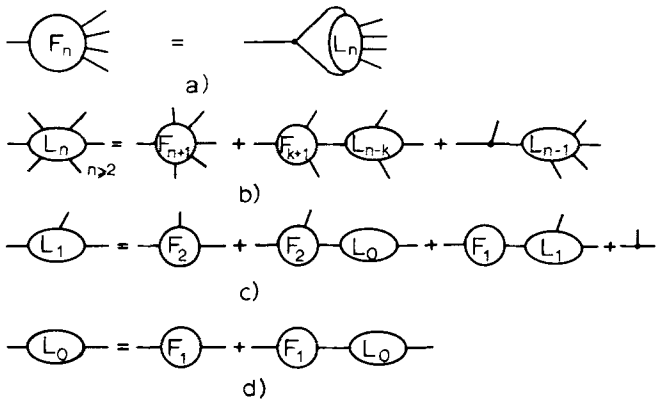

FIG. 1. Graphical relations between the vertex function F_n which is the sum of all one-particle irreducible diagrams and the function L_n which is the sum of all connected diagrams.

infinite set of coupled equations for φ^3 theory from the Schwinger equation in functional derivatives. Appendix B contains the derivation of some inequatities for the integrals used in Sec. 3.

2. DERIVATION OF THE SET OF COUPLED NONLINEAR INTEGRAL EQUATIONS FOR THE GREEN FUNCTIONS OF THE φ^3 THEORY

Let us consider the model described by the interaction Lagrangian

$$\mathscr{L}_{\rm int} = \frac{\lambda}{3}\varphi^3,$$

where λ is the coupling constant, φ-scalar field with mass m.

Usually, the coupled system of equations is derived analytically using variational derivatives with respect to external sources.[11,12] Here, we give a transparent derivation of these equations based on the properties of Feynman diagrams.[13]

The vertex function $F_n(x_0,...,x_n)$ is the sum of all one-particle irreducible diagrams with $n+1$ external lines. To obtain the coupled system of equations for F_n, we introduce the functions L_n. The function $L_n(x,y;z_1,...,z_n)$ is the sum of all connected diagrams with the following property: if by removing one line the diagram becomes disconnected, then the vertices x and y appear in different parts of the diagram.

One can see that in φ^3 theory each one-particle irreducible diagram can be built out of diagrams of sort L, as shown in Fig. 1a. Conversely, each diagram of type L can be built out of one-particle irreducible diagrams and diagrams of type L in one of the ways depicted in Figs. 1b, 1c, and 1d.

These relations among the diagrams bring to the coupled system of equations for the functions $\tilde{F}_n$ and $\tilde{L}_n$ defined as

$$\tilde{F}_n(q_1,...,q_n) = \int \exp\Big(i\sum_{i=1}^{n} q_i x_i\Big) F_n(0,x_1,...,x_n)\,dx_1\cdots dx_n,$$

$$\tilde{L}_n(p;q_1,...,q_n) = \int \exp\Big[i\Big(\sum_{i=1}^{n} q_i x_i + py\Big)\Big] L_n(0,y;x_1,...,x_n) \times dy\,dx_1\cdots dx_n.$$

As seen from Figs. 1b, 1c, and 1d, it is convenient to introduce for all $n\geqslant 0$ the function B_n

$$B_n(p;q_1,...,q_n) = \sum_{k=0}^{n} \rho\Big(\frac{q_1,...,q_k}{q_{k+1},...,q_n}\Big) \times \frac{\tilde{F}_{k+1}(-p-\Sigma_{i=1}^{k} q_i, q_1,...,q_k)}{[(p+\Sigma_{i=1}^{k} q_i)^2+m^2]} \times \tilde{L}_{n-k}\Big(p+\sum_{i=1}^{k} q_i; q_{k+1},...,q_n\Big), \quad (2.1)$$

and for $n\geqslant 1$, $1\leqslant k\leqslant n$, to introduce D_{nk}

$$D_{nk}(p;q_1,...,q_n) = \frac{2\lambda}{(p+q_k)^2+m^2} \times \tilde{L}_{n-1}(p+q_k;q_1,...,q_{k-1};q_{k+1},...,q_n). \quad (2.2)$$

Here $\rho(q_1,...,q_n/q_{k+1},...,q_n)$ means summation over all the divisions of variables $q_1,...,q_n$ into two groups with k and $(n-k)$ elements. Graphical representations corresponding to B_n and D_{nk} are shown in Figs. 2.

Now the relations given in Figs. 1a– 1d among the diagrams of the φ^3 model can be presented in the form of following equations:

$$\tilde{L}_n(p;q_1,...,q_n) = \tilde{F}_{n+1}\Big(-p-\sum_{i=1}^{n} q_i, q_1,...,q_n\Big) + B_n(p;q_1,...,q_n) + \sum_{k=1}^{n} D_{nk}(p;q_1,...,q_n) \quad (2.3)$$

for $n\geqslant 2$, and

$$\tilde{L}_1(p;q) = \tilde{F}_2(-p-q,q) + B_1(p;q) + D_{11}(p;q) + 2\lambda, \quad (2.4)$$

$$\tilde{L}_0(p) = \tilde{F}_1(p) + B_0(p), \quad (2.5)$$

$$\tilde{F}_n(q_1,...,q_n) = \frac{\lambda}{(2\pi)^4}\int \frac{d^4p}{(p^2+m^2)[(p+\Sigma_{i=1}^{n} q_i)^2+m^2]} \times \Big[\tilde{F}_{n+1}\Big(-p-\sum_{i=1}^{n} q_i, q_1,...,q_n\Big) + B_n(p;q_1,...,q_n) + \sum_{k=1}^{n} D_{nk}(p;q_1,...,q_n)\Big] \quad (2.6)$$

for $n\geqslant 2$, and

$$\tilde{F}_1(q) = \frac{\lambda}{(2\pi)^4}\int \frac{d^4p}{(p^2+m^2)[(p+q)^2+m^2]} \times \{B_1(p;q) + D_{11}(p;q)\} + \frac{2\lambda^2}{(2\pi)^4}\int \frac{d^4p}{p^2+m^2}$$

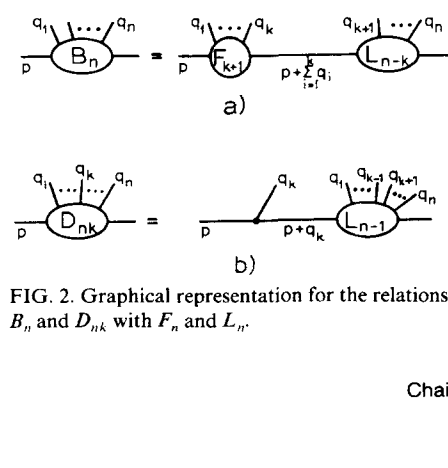

FIG. 2. Graphical representation for the relations of introduced functions B_n and D_{nk} with F_n and L_n.

$$\times\left[\frac{1}{(p+q)^2+m^2}-\frac{1}{p^2+m^2}\right]+\mu(\lambda). \quad (2.7)$$

The function $\tilde{F}_1$ coincides with the mass operator. Thus in Eq. (2.7) subtraction is performed. As seen, the introduction of functions B_n and D_{nk} allowed us to separate in Eq. (2.7) the term corresponding to the only divergent diagram of φ^3 theory. The term $\mu(\lambda)$ corresponds to the finite renormalization of mass.

Let us recall that we are considering the functions $\tilde{F}_n$, $\tilde{L}_n$, B_n, and D_{nk} in the Euclidean region of the variables. The continuation to Euclidean region is done by means of a Wick rotation.[14] Thus the internal propagators $1/(p^2+m^2)$, which enter our expressions, do not have singularities.

3. PRINCIPLE OF CONTRACTION MAPPING (OR BANACH'S FIXED POINT THEOREM) AND ITS USE FOR OUR PROBLEM

The system of equations (2.1)–(2.7) can be written in the form

$$\varphi = A\varphi,$$

where φ is the set of functions $\tilde{F}_n$, $\tilde{L}_n$, B_n, D_{nk} and A is a nonlinear operator in the space of functions φ. This allows us to investigate the question of existence and uniqueness of the solution of system (2.1)–(2.7) with the help of the contraction mapping principle or Banach's fixed point theorem (see, e.g., Ref. 1). According to this principle, the equation $A\varphi = \varphi$ has a unique solution inside a sphere of radius R in the Banach space $\|\varphi\| \leqslant R$, if there exists a real number κ with $0 \leqslant \kappa < 1$, such that for all the elements φ_1 and φ_2 of the Banach space (i.e., $\|\varphi_1\| \leqslant R$ and $\|\varphi_2\| \leqslant R$) the following conditions are fulfilled:

$$\|A\varphi_1\| \leqslant R, \quad \|A\varphi_2\| \leqslant R,$$
$$\|A\varphi_1 - A\varphi_2\| \leqslant \kappa\|\varphi_1 - \varphi_2\|.$$

In order to use the contraction mapping principle to the system (2.1)–(2.7), let us introduce our norm as[6]

$$\|\varphi\| = \max\left\{\sup_n \|\tilde{F}_n\|, \sup_n \|\tilde{L}_n\|, \sup_n \|B_n\|, \sup_{n,k} \|D_{nk}\|\right\},$$

where

$$\|\tilde{F}_n\| = f_n^{-1}\sup\left|\frac{1}{m^3}\left(\left|\sum_{i=1}^{n} q_i\right| + m\right)\tilde{F}_n(q_1,\ldots,q_n)\right| \quad \text{for } n \geqslant 2,$$

$$\|\tilde{F}_1\| = f_1^{-1}\sup\left|\frac{|q_1|+m}{m(q_1^2+m^2)}\tilde{F}_1(q_1)\right|,$$

$$\|\tilde{L}_n\| = l_n^{-1}\sup\left|\frac{1}{m^2}\tilde{L}_n(p;q_1,\ldots,q_n)\right| \quad \text{for } n \geqslant 1,$$

$$\|\tilde{L}_0\| = l_0^{-1}\sup\left|\frac{|p|+m}{m(p^2+m^2)}\tilde{L}_0(p)\right|,$$

$$\|B_n\| = b_n^{-1}\sup\left|\frac{|p|+m}{m^3}B_n(p;q_1,\ldots,q_n)\right| \quad \text{for } n \geqslant 1,$$

$$\|B_0\| = b_0^{-1}\sup\left|\frac{1}{m^2}B_0(p)\right|,$$

$$\|D_{nk}\| = d_n^{-1}\sup\left|\frac{(p+q_k)^2+m^2}{m^4}D_{nk}(p;q_1,\ldots,q_n)\right| \quad \text{for } n \geqslant 2,$$

and

$$\|D_{11}\| = d_1^{-1}\sup\left|\frac{|p+q_1|}{m^3}D_{11}(p;q_1)\right|,$$

with f_n, l_n, b_n, and d_n some normalization constants (real and positive) to be defined later and $|q| \equiv (q^2)^{1/2}$.

Using the inequalities

$$\left|\int \frac{d^4p}{p^2+m^2}\left[\frac{1}{(p+q)^2+m^2}-\frac{1}{p^2+m^2}\right]\right| \leqslant \alpha\frac{|q|}{m}, \quad (3.1)$$

$$\int \frac{d^4l}{(l^2+m^2)[(l-k)^2+m^2][(l-p)^2+m^2]} \leqslant \frac{\beta}{m(m+|k|)}, \quad (3.2)$$

$$\int \frac{d^4l}{(l^2+m^2)[(l-k)^2+m^2](|l-k|+m)} \leqslant \frac{\gamma}{m+|k|}, \quad (3.3)$$

the derivation of which is given in Appendix B, we obtain the following sufficient conditions for the existence and uniqueness of the solution of the system (2.1)–(2.7) in the region $\|\varphi\| \leqslant R$:

$$2R\sum_{k=0}^{n} C_n^k f_{k+1} l_{n-k} \leqslant \kappa b_n \quad \text{for } n \geqslant 0, \quad (3.4)$$

where $C_n^k = n!/k!(n-k)!$,

$$\frac{2\lambda}{m^2} l_{n-1} \leqslant \kappa d_n \quad \text{for } n \geqslant 1, \quad (3.5)$$

$$f_{n+1} + b_n + nd_n \leqslant \kappa l_n \quad \text{for } n \geqslant 2, \quad (3.6)$$

$$f_1 + b_1 + d_1 + \frac{2\lambda}{m^2 R} \leqslant l_1, \quad (3.7)$$

$$f_1 + 2b_0 \leqslant \kappa l_0, \quad (3.8)$$

$$\frac{\lambda}{(2\pi)^4}[(f_{n+1}+b_n)\gamma + nd_n\beta] \leqslant \kappa f_n \quad \text{for } n \geqslant 2, \quad (3.9)$$

$$\frac{\lambda\gamma}{(2\pi)^4}(b_1+d_1) + \frac{4\lambda^2\alpha}{(2\pi)^4 Rm^2} + \frac{2|\mu(\lambda)|}{Rm^2} \leqslant f_1. \quad (3.10)$$

4. DERIVATION OF FINAL RESULTS

Since the normalization factors f_n, l_n, b_n, and d_n, introduced through the definition of norms in Sec. 3, can be chosen arbitrarily, the system of inequalities (3.4)–(3.10) can be considered as equations with respect to the unknowns f_n, l_n, b_n, and d_n. In this way, the use of contraction mapping principle to the system (2.1)–(2.7) has brought us to the question of existence of solution for the system (3.4)–(3.10) in the region $f_n > 0$, $l_n > 0$, $b_n > 0$, $d_n > 0$, and $R > 0$.

Let us first consider the truncated chain of equations

(2.1)–(2.7), i.e., consider only those equations, in the left hand side of which are contained functions with the number n such that $n \leqslant N$. The functions $\widetilde{F}_{N+1}$ which appear in the right-hand side of these equations we put equal to zero. Thus, the truncated system has solutions if the system (3.4)–(3.10) is compatible for $n \leqslant N$. We show that for sufficiently small values of λ this is the case. For this matter we shall assume that $\mu(\lambda) \to 0$ for $\lambda \to 0$. This assumption can be done, since in the perturbation theory $\mu(\lambda) \simeq c\lambda^2$ in the neighborhood of $\lambda = 0$. Notice, however, that one can in general drop the term $\mu(\lambda)$. In that case the vertex functions would correspond to a field with the renormalized mass $m + \delta m$.

So, let R and l_n $(0 \leqslant n \leqslant N)$ be arbitrary real positive numbers. Then one can choose b_n and d_n $(0 \leqslant n \leqslant N)$ such that

$$b_n + n d_n \leqslant \tfrac{1}{2}\kappa l_n \quad \text{for } n \geqslant 1,$$

$$2b_0 \leqslant \tfrac{1}{2}\kappa\, l_0.$$

Now it is possible to choose f_n $(1 \leqslant n \leqslant N)$ sufficiently small in order that inequalities (3.4), (3.6) and (3.8) are fulfilled and also $f_2 < \frac{1}{4} l_1$. In the left-hand side of inequalities (3.5), (3.9), and (3.10) there stands the factor λ. Since we are working with the truncated system of equations, then (3.5), (3.9), and (3.10) form a finite system of inequalities. Therefore, for sufficiently small λ these inequalities will be fulfilled. The inequality (3.7) is also satisfied, since $f_2 + b_1 + d_1 < \frac{3}{4} l_1$. Thus we have shown that the truncated system of equations (2.1)–(2.7) has a solution for sufficiently small values of λ.

However, the introduced method does not allow us to prove the existence of solution of the complete system (2.1)–(2.7) for $\lambda \neq 0$. We can show that the system (3.4)–(3.10) becomes incompatible for $N = \infty$.

Indeed, from ineq. (3.9) we have

$$[\lambda\gamma/(2\pi)^4] b_n < f_n, \tag{4.1}$$

$$[\lambda\gamma/(2\pi)^4] f_n < f_{n-1} \tag{4.2}$$

since $0 \leqslant \kappa < 1$.

From (3.4) we have

$$Rn(n-1) f_{n-1} l_2 < b_n. \tag{4.3}$$

Combining these inequalities, we obtain

$$Rn(n-1) l_2 [\lambda\gamma/(2\pi)^4]^2 b_n < b_n, \tag{4.4}$$

from which for all $n \geqslant 2$

$$Rn(n-1) l_2 [\lambda\gamma/(2\pi)^4]^2 < 1. \tag{4.5}$$

This is possible only for $\lambda = 0$.

We have demonstrated that the system (3.4)–(3.10) with $N = \infty$ is incompatible for the values of coupling constant $\lambda \neq 0$. Thus the principle of contraction mapping as used in our approach, allows for the existence of solution of system of equations (2.1)–(2.7) for the Green functions of $(\lambda/3)\varphi^3$ theory only for $\lambda = 0$.

As we have already mentioned in the Introduction, the existence of solutions for the Green functions in quantum field theory is of great importance. The use of a powerful mathematical tool of Banach's fixed point theorem gave us, however, a unsatisfactory answer, as mentioned above. Possible (although not necessarily so) reasons for that may be the estimates of inequalities and majorizations and/or the definition of norms used for the Green functions. A more detailed analysis of these majorizations and introduction of other norms and, as well, other fixed point approaches like topological degree[3-5] are therefore quite necessary and might lead to spectacular results. It is known that field theories like $(\lambda/3)\varphi^3$ have zero radius of convergence in λ. So it could appear very hard to use the contaction mapping method, however the norms are altered, to make such an iterative solution well defined. Thus it would be very interesting to look for an argument (such as a counter example) in order to rule out the possible expectation that contraction mapping will always fail when there is zero radius of convergence of the iterative solution. For other quantum field theory models like QED with an infinite number of divergent graphs the situation is more complicated. We have chosen φ^3 theory because of presence of only one divergent diagram which simplifies the analysis.

ACKNOWLEDGMENT

The authors wish to express their gratitude to N.N. Bogoliubov and C. Cronström for valuable discussions.

APPENDIX A: FUNCTIONAL EQUATIONS FOR THE GREEN FUNCTIONS AND EQUIVALENT INFINITE SYSTEM OF EQUATIONS

The field theory with φ^n interaction: Consider the interacting meson field described by the Lagrangian

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_{\text{int}} \tag{A1}$$

where

$$\mathscr{L}_0 = \tfrac{1}{2}[(\partial\varphi/\partial x)^2 - m^2\varphi^2]$$

and

$$\mathscr{L}_{\text{int}} = (g/n)\,\varphi^n \tag{A2}$$

is the interaction Lagrangian. All the operators here and further on are taken in the interaction representation.

Lagrangian (A1) corresponds to the equation of motion

$$(\Box + m^2)\varphi(x) = j(x) = g\varphi^{n-1}(x) \tag{A3}$$

and to the scattering matrix

$$S = T\exp\left[i\int j(x)\varphi(x)\,dx\right], \tag{A4}$$

where T is the operator of chronological ordering. By definition, Green functions are called the vacuum expectation of the chronological product of operators. In particular, the one-particle Green function is given as

$$G(x,y) = (1/S_0)\langle 0|\, T\varphi(x)\varphi(y) S\,|0\rangle, \tag{A5}$$

where $S_0 = \langle 0|S|0\rangle$.

Let us derive now functional equation for the Green function $G(x,y)$. For this matter, introduce an auxiliary source which corresponds to a current $J(x)$ and interaction energy $J(x)\varphi(x)$. Then (A3) and (A4) will take the following forms:

$$(\Box + m^2)\varphi(x) = g\varphi^{n-1}(x) + J(x), \tag{A6}$$

$$S = T\exp\left\{i\int [j(x) + J(x)]\varphi(x)\,dx\right\}. \tag{A7}$$

As seen, the Green functions become functionals of $J(x)$, and one can write for them equations with functional derivatives. Multiplying by $(1/S_0)\,\varphi(x)S$ both sides of (A.6) from right side, acting with the operator T, and taking the vacuum expectation, we get

$$\frac{1}{S_0}\langle 0|T(\Box+m^2)\varphi(x)\varphi(y)S|0\rangle = \frac{1}{S_0}\langle 0|T[g\varphi^{n-1}(x)+J(x)]\varphi(y)S|0\rangle. \tag{A8}$$

Operators $\Box$ and T do not commute and one has for them

$$T\left[\frac{\partial}{\partial x_0^2}\varphi(x)\varphi(y)\right] = \frac{\partial}{\partial x_0^2}[T\varphi(x)\varphi(y)] - \delta(x-y). \tag{A9}$$

From (A5), (A9), and (A8) one gets

$$(\Box+m^2)G(x,y) - \frac{1}{S_0}g\langle 0|T\varphi^{n-1}(x)\varphi(y)S|0\rangle - J(x)\langle\varphi(y)\rangle = -i\delta(x-y), \tag{A10}$$

where

$$\langle\varphi(x)\rangle = \frac{1}{S_0}\langle 0|\varphi(x)S|0\rangle. \tag{A11}$$

From the relation

$$\frac{\delta S}{\delta J(x)} = iT[\varphi(x)S], \tag{A12}$$

which follows from (A7), one has

$$\langle 0|T\varphi^{n-1}(x)\varphi(y)S|0\rangle = \frac{1}{i^{n-2}}\frac{\delta^{n-2}}{\delta J^{n-2}(x)}\langle 0|T\varphi(x)\varphi(y)S|0\rangle = \frac{1}{i^{n-2}}\frac{\delta^{n-2}}{\delta J^{n-2}(x)}[G(x,y)S_0]. \tag{A13}$$

Thus, finally one gets the equation for the Green function as

$$(\Box+m^2)G(x,y) - \frac{g}{i^{n-2}S_0}\frac{\delta^{n-2}}{\delta J^{n-2}(x)}[G(x,y)S_0] - J(x)\langle\varphi(y)\rangle = -i\delta(x-y). \tag{A14}$$

Consider now the φ^3 interaction, i.e., the case $n=3$, and for further consideration let us introduce a function $\Delta(x,y)$ equal to functional derivative of $\langle\varphi(x)\rangle$:

$$\frac{\delta\langle\varphi(x)\rangle}{\delta J(y)} \equiv i\Delta(x,y) = iG(x,y) - i\langle\varphi(x)\rangle\langle\varphi(y)\rangle. \tag{A15}$$

When $J(x)=0$, the function $G(x,y)$ and $\Delta(x,y)$ coincide. Replacing G by Δ, Eq. (A14) becomes

$$\left[\Box+m^2+ig\frac{\delta}{\delta J(x)} - 2g\langle\varphi(x)\rangle\right]\Delta(x,y) = -i\delta(x-y). \tag{A16}$$

Notice also that the functional derivative of the Green function $G(x,y)$ is equal to

$$\frac{\delta G(x,y)}{\delta J(x)} = \frac{i}{S_0}\langle 0|T\varphi(x)\varphi(y)\varphi(x)S|0\rangle - \frac{1}{S_0^2}\langle 0|T\varphi(x)\varphi(y)S|0\rangle\langle 0|T\varphi(x)S|0\rangle = iK(x,y;x) - G(x,y)\langle\varphi(x)\rangle. \tag{A17}$$

Infinite system of coupled equations: By substituting (A17) into (A14), one has eliminated the functional derivative and has arrived at a differential equation. But then there appears a new unknown function $K(x,y;x)$, for which one needs an additional equation. This can be achieved by taking the functional derivative of (A10) with respect to $J(x)$ and using (A12). Then in the equation there appears a new function $K(x,y;x,x)$. Continuing this procedure, we arrive at an infinite system of differential equations for Green functions. Thus the Schwinger equation in functional derivatives is equivalent to an infinite system of coupled differential equations.

Let us find this system explicitly. Multiplying Eq. (A6) by $S_0^{-1}S$ from the right side and taking the vacuum expectation gives

$$(\Box m^2)\langle\varphi(x)\rangle = gG(x,x) + J(x) = g\Delta(x,x) + g\langle\varphi(x)\rangle^2 + J(x). \tag{A18}$$

Further, it is more convenient to use Eq. (A18) in momentum space, i.e., its Fourier transform. Then (A18) becomes

$$(k^2-m^2)\frac{1}{(2\pi)^4}\int\langle\varphi(k)\rangle e^{ikx}\,dk = g\frac{1}{(2\pi)^4}\int\Delta(k,p)e^{i(k-p)}\,dk\,dp + g\frac{1}{(2\pi)^8}\int\langle\varphi(k)\rangle e^{ikx}\langle\varphi(p)\rangle e^{ipx}\,dk\,dp + \int J(k)e^{ikx}\,dk. \tag{A19}$$

Finally, Eq. (A18) in momentum space is given by

$$(k^2-m^2)\langle\varphi(k)\rangle = -g\int\Delta(p+k,p)\,dp - \frac{g}{(2\pi)^4}\int\langle\varphi(k-p)\rangle\langle\varphi(p)\rangle\,dp - J(k). \tag{A20}$$

By acting on both sides of Eq. (A20) the nth functional derivative with respect to $J(s)$ (and using the Leibnitz formula) we obtain

$$(k^2-m^2)\langle\varphi^n(k,s_1,\ldots,s_n)\rangle = -g\int\Delta^n(p+k,p,s_1,\ldots,s_n)dp - g\frac{1}{(2\pi)^4}\sum_{m=0}^{n}C_n^m\int dp\langle\varphi^m(k-p,s_1,\ldots,s_m)\rangle \times\langle\varphi^{n-m}(p,s_{m+1},\ldots,s_n)\rangle - \delta_m\delta\Big(k-\sum_{i=1}^{n}s_i\Big), \tag{A21}$$

where

$$\Delta^n(p,q,s_1,\ldots,s_n) = \frac{\delta^n\Delta(p,q)}{\delta J(s_1)\cdots\delta J(s_n)}, \tag{A22}$$

$$\langle\varphi^n(k,s_1,\ldots,s_n)\rangle = \frac{\delta^n\langle\varphi(k)\rangle}{\delta J(s_1)\cdots\delta J(s_n)}, \tag{A23}$$

and $C_n^m = n!/m!(n-m)!$ are the binomial coefficients. For $J(x)=0$ from energy–momentum conservation one can write (A22) and (A23) in the form

$$\Delta^n(p,q,s_1,\ldots,s_n) = \Delta^n(p,s_1,\ldots,s_n)\delta(p-q-s_1\cdots-s_n), \tag{A24}$$

$$\langle\varphi^n(k,s_1,\ldots,s_n)\rangle = D^{n-1}(k,s_1,\ldots,s_n)\delta(k-s_1\cdots-s_n). \tag{A25}$$

Substitution of (A24) and (A25) into (A21) gives then

$$\begin{aligned}(k^2-m^2)D^n(k,s_1,\ldots,s_n) &= -\delta_{n0} - g\int \Delta^{n+1}\Big(k+p,s_1,\ldots,s_n,k-\sum_{i=1}^{n}s_i\Big)\,dp \\ &\quad - \frac{g}{(2\pi)^4}\sum_{m=0}^{n+1} C_{n+1}^{m} D^{m-1}\Big(\sum_{i=1}^{m} s_i,s_1,\ldots,s_{m-1}\Big) \\ &\quad \times D^{n-m}\Big(k-\sum_{i=1}^{m}s_i,s_{m+1},\ldots,s_n\Big).\end{aligned} \tag{A26}$$

From (A15), (A22), and (A23)–(A25) follows the relation

$$i\Delta^n(p,s_1,\ldots,s_n) = D^n(p,s_1,\ldots,s_n). \tag{A27}$$

Finally, the infinite system of coupled equations for the functions $\Delta^n(k,s_1,\ldots,s_n)$ for the case of $(g/3)\varphi^3$ interaction is

$$\begin{aligned}(k^2-m^2)\Delta^n(k,s_1,\ldots,s_m) &= i\delta_{n0} + ig\int dp\Delta^{n+1}\Big(k+p,s_1,\ldots,s_n,k-\sum_{i=1}^{n}s_i\Big) \\ &\quad - i\frac{g}{(2\pi)^4}\sum_{m=0}^{n+1} C_{n+1}^{m}\Delta^{m-1}\Big(\sum_{i=1}^{m}s_i,s_1,\ldots,s_{m-1}\Big) \\ &\quad \times\Delta^{n-m}\Big(k-\sum_{i=1}^{m}s_i,s_{m+1},\ldots,s_n\Big).\end{aligned} \tag{A28}$$

Thus we have arrived to the coupled system of nonlinear integral equations which can be schematically written as: $\Delta = A\Delta$.

APPENDIX B: DERIVATION OF INEQUALITIES (3.1)–(3.3)

Inequality (3.1): Going to "$-\alpha$"-representation, we write the left-hand side of (3.1) in the form

$$\begin{aligned}&\int d^4p\int_0^\infty\int_0^\infty d\alpha\,d\beta\,e^{-\alpha(p^2+m^2)}\left(e^{-\beta[(p+q)^2+m^2)]}-e^{-\beta(p^2+m^2)}\right) \\ &\quad = \int_0^\infty\int_0^\infty d\alpha\,d\beta\,e^{-(\alpha+\beta)m^2}\left(e^{-\beta q^2}\int d^4p e^{-(\alpha+\beta)p^2-2\beta pq}-\int d^4p e^{-(\alpha+\beta)p^2}\right).\end{aligned} \tag{B1}$$

With the help of the formula[11]

$$\int d^4p\,e^{-\xi p^2-2qp} = (\pi^2/\xi^2)e^{q^2/\xi}, \tag{B2}$$

we perform integration in (B1) with respect to p. Then (B1) can be written as

$$\int_0^\infty\int_0^\infty d\alpha\,d\beta\,e^{-(\alpha+\beta)m^2}[\pi^2/(\alpha+\beta)^2]\left[e^{-q^2[\beta-\beta^2/(\alpha+\beta)]}-1\right]. \tag{B3}$$

Introducing new variables

$\beta=\lambda\xi$, $\alpha=\lambda(1-\xi)$, with $\partial(\alpha,\beta)/\partial(\lambda,\xi)=\lambda$,

(B3) becomes

$$\begin{aligned}\pi^2\int_0^1 d\xi\int_0^\infty\frac{d\lambda}{\lambda}e^{-\lambda m^2}(e^{-\lambda q^2(\xi-\xi^2)}-1) &= \pi^2\int_0^1 d\xi\int_0^\infty d\varphi\int_0^\infty d\lambda\,(e^{-\lambda[\varphi+m^2+q^2(\xi-\xi^2)]}-e^{-\lambda(\varphi+m^2)}) \\ &= \pi^2\int_0^1 d\xi\int_0^\infty d\varphi\left[\frac{1}{\varphi+m^2+q^2(\xi-\xi^2)}-\frac{1}{\varphi+m^2}\right] \\ &= -\pi^2\int_0^1 d\xi\ln\left[1+\frac{q^2}{m^2}(\xi-\xi^2)\right].\end{aligned} \tag{B4}$$

The absolute magnitude of this integral is less than

$$\pi^2\ln(1+q^2/4m^2)\leqslant\pi^2\ln(1+|q|/2m)^2\leqslant\pi^2|q|/m, \tag{B5}$$

which is the inequality (3.1).

Inequality (3.2): We write the left-hand side of (3.2) in "α"-representation and perform integration over l using (B2). Then the left-hand side of (3.2) can be written as

$$\begin{aligned}&\int_0^\infty\int_0^\infty\int_0^\infty d\alpha_1\,d\alpha_2\,d\alpha_3\int d^4l\exp\{-\alpha_1(l^2+m^2)-\alpha_2[(l-k)^2+m^2]-\alpha_3[(l-p)^2+m^2]\} \\ &= \int_0^\infty\int_0^\infty\int_0^\infty d\alpha_1\,d\alpha_2\,d\alpha_3\exp[-m^2(\alpha_1+\alpha_2+\alpha_3)-\alpha_2k^2-\alpha_3p^2]\int d^4l\exp[-(\alpha_1+\alpha_2+\alpha_3)l^2+2l(\alpha_2k+\alpha_3p)] \\ &= \pi^2\int_0^\infty\int_0^\infty\int_0^\infty d\alpha_1\,d\alpha_2\,d\alpha_3\exp[-m^2(\alpha_1+\alpha_2+\alpha_3)-\alpha_2k^2-\alpha_3p^2]\frac{\exp[(\alpha_2k+\alpha_3p)^2/(\alpha_1+\alpha_2+\alpha_3)]}{(\alpha_1+\alpha_2+\alpha_3)^2}.\end{aligned} \tag{B6}$$

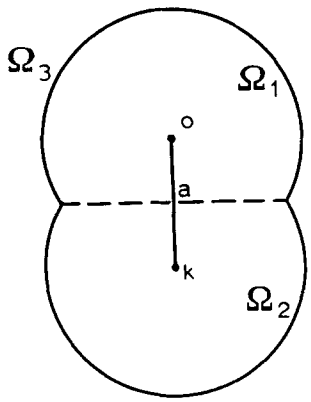

FIG. 3. The division of four-dimensional integration space into three regions Ω_1, Ω_2, and Ω_3.

Introducing new variables

$$\alpha_1 = \lambda(1-\xi_1-\xi_2), \quad \alpha_2 = \lambda\xi_1, \quad \alpha_3 = \lambda\xi_2 \quad \text{with} \quad \partial(\alpha_1,\alpha_2,\alpha_3)/\partial(\lambda,\xi_1,\xi_2) = \lambda^2,$$

we see that (B6) becomes equal to

$$\pi^2 \int_{\substack{\xi_1,\xi_2>0 \\ \xi_1+\xi_2<1}} d\xi_1\, d\xi_2 \int_0^\infty d\lambda \exp\{-\lambda[m^2+\xi_1 k^2+\xi_2 p^2-(\xi_1 k+\xi_2 p)^2]\}$$
$$= \pi^2 \int_{\substack{\xi_1,\xi_2>0 \\ \xi_1+\xi_2<1}} d\xi_1\, d\xi_2\, [m^2+\xi_1 k^2+\xi_2 p^2-(\xi_1 k+\xi_2 p)^2]^{-1} \tag{B7}$$

For fixed k, the integrand in (B7) reaches its maximum at the point

$$p = [\xi_1/(1-\xi_2)]k$$

with the value of integrand equal to

$$\frac{1}{m^2+k^2\xi_1(1-\xi_1-\xi_2)/(1-\xi_2)}.$$

Therefore, what we need is to prove the following inequality:

$$\int_{\substack{\xi_1,\xi_2>0 \\ \xi_1+\xi_2<1}} d\xi_1\, d\xi_2 \frac{1}{m^2+k^2\xi_1(1-\xi_1-\xi_2)/(1-\xi_2)} \leqslant \frac{\beta/\pi^2}{m(m+|k|)}. \tag{B8}$$

For this purpose we divide the integration region in (B8) into two regions:

$$S_1: \quad \delta \leqslant \xi_1, \quad 0 \leqslant \xi_2, \quad \xi_1+\xi_2 \leqslant 1-\delta, \qquad S_2: \quad 0 \leqslant \xi_1 \leqslant \delta, \quad 0 \leqslant \xi_2, \quad 1-\delta \leqslant \xi_1+\xi_2 \leqslant 1, \tag{B9}$$

where

$$\delta = \frac{1}{1+|k|/m}.$$

The area of region S_2 is less than 2δ. Thus the integral over this region is less than

$$2\delta/m^2 = 2/m(m+|k|). \tag{B10}$$

In the region S_1, we have

$$\xi_1(1-\xi_1-\xi_2)/(1-\xi_2) \geqslant \delta/2.$$

The area of this region is less than $\frac{1}{2}$; therefore, the value of integral in this region is less than

$$\frac{1}{2}\frac{1}{m^2+k^2\delta/2} = \left[m^2\left(2\frac{k^2/m^2}{1+|k|/m}\right)\right]^{-1} \leqslant \frac{1}{m(m+|k|)}. \tag{B11}$$

Combining (B10) and (B11), one obtains the inequality (3.2) with $\beta = 3\pi^2$.

Inequality (3.3): By scaling the variables to l/m and k/m it is obvious that it is enough to prove the inequality (3.3) only for $m = 1$. Let us divide the four-dimensional space of integration into three regions Ω_1, Ω_2, and Ω_3 and prove that the estimate (3.3) is valid for each region. Let σ_1 and σ_2 be two subspaces ($\sigma_1 \cup \sigma_2 = R^4$, $0 \in \sigma_1$, $k \in \sigma_2$) divided by a hypersurface which passes through $\bar{a} = \frac{1}{2}\bar{k}$ perpendicular to $\bar{k}$. Let S_1 and S_2 be spheres of radius $|\bar{k}|$ with their centers at $\bar{0}$ and $\bar{k}$ respectively. We take $\Omega_1 = S_1 \cap \sigma_1$, $\Omega_2 = S_2 \cap \sigma_2$, and $\Omega_3 = \overline{\Omega_1 \cup \Omega_2}$ (Fig. 3).

In region Ω_1 we have

$$\frac{1}{(l^2+1)[(l-k)^2+1](|l-k|+1)} \leqslant \frac{1}{l^2+1}\,\frac{1}{(\frac{1}{4}k^2+1)(\frac{1}{2}|k|+1)}.$$

Therefore, the integral I_1 over the region Ω_1 is

$$I_1 \leqslant \frac{1}{(\frac{1}{4}k^2+1)(\frac{1}{2}|k|+1)}\int_{\Omega_1}\frac{d^4l}{l^2+1} \leqslant \frac{1}{(\frac{1}{4}k^2+1)(\frac{1}{2}|k|+1)}\int_0^{|k|}\frac{\epsilon_0 r^3\,dr}{r^2+m^2} \leqslant \frac{\frac{1}{2}\epsilon_0 k^2}{(\frac{1}{4}k^2+1)(\frac{1}{2}|k|+1)} \leqslant \frac{2\epsilon_0}{\frac{1}{2}|k|+1}$$
$$\leqslant \frac{4\epsilon_0}{|k|+1}, \tag{B12}$$

where ϵ_0 is the area of the surface of the unit four-dimensional sphere.

In the region Ω_2 we have

$$\frac{1}{(l^2+1)[(l-k)^2+1](|l-k|+1)} \leqslant \frac{1}{(\frac{1}{4}k^2+1)}\,\frac{1}{[(l-k)^2+1](|l-k|+1)}.$$

Then the integral I_2 over the region Ω_2 is

$$I_2 \leqslant \frac{1}{(\frac{1}{4}k^2+1)}\int_{\Omega_2}\frac{d^4l}{[(l-k)^2+1](|l-k|+1)} \leqslant \frac{1}{(\frac{1}{4}k^2+1)}\int_0^{|k|}\frac{\epsilon_0 r^3\,dr}{(r^2+1)(r+1)} \leqslant \frac{\epsilon_0|k|}{\frac{1}{4}k^2+1} \leqslant \frac{8\epsilon_0}{|k|+1} \tag{B13}$$

In the region Ω_3 we have $|l| \geqslant |l-a| - |a|$ and $|l| \geqslant 2|a|$. Therefore, $|l| \geqslant |l-a|$. Analogously, in this region $|l-k| \geqslant \frac{2}{3}|l-a|$. This region is contained in the complementary to the sphere of radius $R = (3^{1/2}/2)|k|$. Thus the integral I_3 over the region Ω_3 is

$$I_3 \leqslant \int_R^\infty \frac{\epsilon_0 r^3\,dr}{(\frac{4}{9}r^2+1)^2(\frac{2}{3}r+1)} \leqslant \frac{27}{8}\epsilon_0\int_R^\infty \frac{dr}{\frac{4}{9}r^2+1} \leqslant \frac{27}{8}\epsilon_0\int_R^\infty \frac{dr}{[(2/\sqrt{3})r+1]^2\frac{1}{6}} \leqslant \frac{81}{4}\,\frac{1}{|k|+1}. \tag{B14}$$

Combining (B12), (B13), and (B14), we are led to inequality (3.3).

[1]L.V. Kantorovich and G.P. Akilov, *Functional Analysis in Normed Space* (Pergamon, Oxford, 1964).

[2]Strictly speaking, the existence of solution in Euclidean region does not automatically lead to existence in Minkowski space. However, the nonexistence in the former space, most probably, implies also the nonexistence in Minkowski space.

[3]Jane Cronin, "Fixed Points and Topological Degree in Nonlinear Analysis," Am. Math. Soc. Math. Surveys, No. 11 (1964).

[4]M.A. Krasnosel'skii, *Topological Methods in the Theory of Nonlinear Integral Equations* (Pergamon, Oxford, 1964).

[5]J.G. Taylor, Proc. Camb. Phil. Soc. **63**, 335 (1967).

[6]N.N. Bogoliubov and V.I. Hazet, Dokl. Akad. Nauk SSSR **66**, 321 (1949); N.N. Bogoliubov, D.J. Petrina, and V.I. Hazet, Theor. & Math. Phys. **1**, 251 (1969).

[7]R.L. Warnock, Phys. Rev. **170**, 1323 (1968); H. McDaniel and R.L. Warnock, Phys. Rev. **180**, 1433 (1969); Nuovo Cimento A **66**, 202 (1970); P.H. Byrnev, V.A. Mesherjakov, and I.P. Nedjalkov, Zh. Eksp. Teor. Fiz. **46**, 663 (1964) [Sov. Phys. JETP **19**, 449 (1964)]; T. Konno and N.F. Nelipa, Theor. & Math. Phys. **12**, 214 (1972) [English transl.: **12**, 769 (1973)].

[8]N.F. Nelipa and G.V. Yakovleva, Proc. of the Triangle Meeting, Bratislava (1975).

[9]D. Atkinson, Nucl. Phys. B **7**, 395 (1968); B **8**, 377 (1968); J. Kupsch, Nucl. Phys. B **11**, 573 (1969); B **12**, 155 (1969).

[10]J.G. Taylor, J. Math. Phys. **7**, 1720 (1966).

[11]J. Schwinger, Proc. Nat. Acad. Sci. **37**, 452, 455 (1951).

[12]N.N. Bogoliubov and D.V. Shirkov, "*Introduction to the Theory of Quantized Fields*" (Interscience, New York, 1959).

[13]Analytical derivation of Schwinger equations for φ^3 interaction and following from them the coupled system of equations are given in Appendix A.

[14]See Ref. 2.

Scalar quantum field in an external gauge field

J. Dimock[a)]

Department of Mathematics, SUNY at Buffalo, Buffalo, New York 14214
(Received 23 October 1978)

Let G be a compact complex Lie group with Lie algebra G' and let $A_\mu \in C_0^\infty(R^4, G')$ be a gauge potential. We construct a gauge covariant local quantum field operator ϕ which solves the equation $[(\partial_\mu + iA_\mu)(\partial^\mu + iA^\mu) + m^2]\phi = 0$ and is a free field in the distant past or future. The associated scattering theory is described by a unitary gauge-invariant S operator.

I. INTRODUCTION

A. A field theory typically consists of a set of fields satisfying some coupled nonlinear differential equations. In the external field problem some of the fields are idealized as fixed functions on space–time (the external fields). The remaining fields then satisfy equations which are typically linear and hence tractable. In this paper we study the case of a scalar field ϕ of mass m in the presence of an external gauge field A_μ. We are mainly interested in the quantum problem where ϕ is a field operator, but begin by discussing the classical case where ϕ is a function.

The starting point is a gauge group G which we take to be a complex compact Lie group. (Real groups could also be considered.) Let $\rho(G)$ be a representation of G by unitary operators on a finite dimension space $\mathbb{C}^m$. Let G' be the Lie algebra for G and let $\rho(G')$ be the representation of G' by skew-Hermitian operators on $\mathbb{C}^m$. The gauge field iA_μ, $0 \leqslant \mu \leqslant 3$, is then a $\rho(G')$-valued function on the space–time R^4. The scalar field ϕ is a $\mathbb{C}^m$-valued function on R^4 and the dynamical equation is

$$\{[\partial_\mu + iA_\mu(x)][\partial^\mu + iA^\mu(x)] + m^2\}\phi(x) = 0. \quad (1.1)$$

Here $x \in R^4$, ∂_μ means $\partial/\partial x_\mu$, and we employ the summation convention $x_\mu y^\mu = x_0 y_0 - \Sigma_{i=1}^3 x_i y_i$. The above equation will be abbreviated as

$$(\Box_A + m^2)\phi = 0. \quad (1.1)'$$

In the special case $G = \rho(G) = U(1), G' = \rho(G') = iR$, the real-valued field A_μ is an electromagnetic potential and the complex-valued ϕ carries an electric charge. Other groups [e.g., SU(2), SU(3),···] correspond to generalized notions of charge (e.g., isospin, strangeness, charm, etc.)

B. The form of the gauge equation is chosen to give a dynamics which is independent of the labeling of charge. This aspect is best understood in the language of fibre bundles, where the coordinate independence is built in. We digress briefly to explain the point. General references are Trautman [1] or Mayer. [2] (The latter also has good references to the physics literature on gauge theories.)

Let P be a principal G bundle over R^4, say $R^4 \times G$, and let $E = P \times_G \mathbb{C}^m$ be the associated vector bundle over R^4 with fibre $\mathbb{C}^m$ and structure group $\rho(G)$. The charged scalar field is now described by a cross section Φ of E, and the external field is described by a connection on P as given by a G'-valued 1-form ω on P.

Coordinates may be introduced by a cross section $x \to (x, g(x))$ of P defined by $g: R^4 \to G$ and called a gauge. One can show that given a gauge g, a cross section Φ determines a function $\phi_g: R_4 \to \mathbb{C}^m$ and a connection ω determines a $\rho(G')$ valued 1-form $iA_g = \Sigma_\mu iA_g^\mu dx_\mu$ on R^4. If g' is another gauge, then it turns out that

$$\phi_{g'} = \gamma^{-1}\phi_g, \quad (1.2)$$

$$iA_{g'}^\mu = \gamma^{-1}(iA_g^\mu)\gamma + \gamma^{-1}\partial^\mu\gamma, \quad (1.3)$$

where $\gamma: R^4 \to \rho(G)$ is defined by

$$\gamma(x) = \rho(g(x)^{-1}g'(x)). \quad (1.4)$$

Conversely, given a family of functions $\{\phi_g\}$ indexed by gauges and satisfying (1.2), there is a unique cross section Φ determining them, and given a family of functions $\{iA_g^\mu\}$ satisfying (1.3) there is a unique connection ω determining them. (Of course, a single element determines the whole family.)

Now we can show that the family of differential operators $\Box_{A_g} + m^2$ for a connection ω defines a (differential) operator $\mathscr{D}_\omega$ on cross sections of E and so is a suitable candidate for defining dynamics. The key point is that the operator is built out of the covariant derivative $\partial^\mu + iA_g^\mu$ and so satisfies

$$(\Box_{A_{g'}} + m^2) = \gamma^{-1}(\Box_{A_g} + m^2)\gamma. \quad (1.5)$$

Now for a cross section Φ as given by the family $\{\phi_g\}$ we define a new family $\psi_g = (\Box_{A_g} + m^2)\phi_g$ and have $\psi_{g'} = \gamma^{-1}\psi_g$ so that $\{\psi_g\}$ defines a cross section Ψ. Thus an operator $\mathscr{D}_\omega: \Phi \to \Psi$ is defined without making a special choice of gauge.

C. For the quantum gauge theory we assume that A_μ has compact support, in fact $A_\mu \in C_0^\infty(R^4, \rho(G'))$. Then we show that there exists a local quantum field operator ϕ satisfying the gauge equation (1.1) in the sense of distributions. The field ϕ is constructed to be equal to a standard free quantum field ϕ_{in} in the distance past. It also reduces to a free field ϕ_{out} in the distant future. We show that there is a unitary operator S connecting these fields by $\phi_{\text{out}} = S^{-1}\phi_{\text{in}}S$.

[a)]Supported by NSF Grant #PHY77-21740.

 0022-2488/79/081791-06$01.00

Then S describes the scattering of the particles created by the field. Finally, we study the effect of a change in gauge on ϕ, S and interpret the result in terms of fibre bundles. These results are all presented in Sec. III. As we shall see, they depend on having sufficiently good control over the classical problem, which we treat in Sec. II.

D. We now discuss some of the related literature. General discussions of the external field problem are given by Wightman [3,4] and Seiler. [5] The case of external gauge fields has apparently not been previously considered. Several authors have considered a scalar field in an external electromagnetic field. However, the unitary result presented here is new even in this case. It was previously only known for $A_1 = A_2 = A_3 = 0$ (Schroer, Seiler, and Swieca [6]) or for A_μ sufficiently small (Belissard [7]).

We also mention the vaguely related work of Schrader [8] on the construction of a Euclidean $\mathscr{P}(\phi)_2$ theory in an external gauge field.

In another paper, [9] the problem of a scalar quantum field in an external gravitational field is treated. This problem is formally quite similar to the gauge problem.

II. THE CLASSICAL PROBLEM

The goal in this section is to construct advanced and retarded fundamental solutions $E_A^\pm$ for the differential operator $(\Box_A + m^2)$, and to obtain certain estimates on $E_A^\pm$ related to the smoothness of the kernel. The construction of the fundamental solutions is not particularly difficult since in leading order we have a diagonal system of strictly hyperbolic differential operators with constant coefficients. Nevertheless, we go into some detail since the smoothness estimates depend on the details of the construction.

The method consists of regarding $\Box_A + m^2$ as an operator on a certain scale of Hilbert spaces consisting of functions which together with a certain number of derivatives vanish exponentially as $x_0 \to -\infty$ or as $x_0 \to +\infty$. With the Cauchy data built into the Hilbert space, one uses *a priori* estimates and a perturbation argument to show that $\Box_A + m^2$ actually has an inverse operator. The technique is essentially that of Leray. [10]

To define the Hilbert spaces, let $\tau \in R$, let $n = (1,0,0,0)$, and let H_τ be the Hilbert space of all measurable functions $f: R^4 \to \mathbb{C}^m$ such that $\|f\|_\tau = (f,f)_\tau^{1/2}$ is finite, where the inner product is

$$(f,g)_\tau = \int \sum_{j=1}^{m} \overline{f_j(x)} g_j(x) e^{-2\tau nx} dx,$$

where $nx = \Sigma_\mu n_\mu x_\mu = x_0$. If $f \in H_\tau$ and $g \in H_{-\tau}$, then we may also form

$$\langle f,g\rangle = \int \sum_{j=1}^{m} f_j(x) g_j(x) dx$$

and have $|\langle f,g\rangle| \leqslant \|f\|_\tau \|g\|_{-\tau}$. Thus every $g \in H_{-\tau}$ defines an element of the dual space H_τ'. In fact, one can identify $H_{-\tau}$ with H_τ' via this pairing.

For each τ and $s = 0,1,2,\cdots$ we next define $H_{\tau,s}$ to be all measurable functions $f: R^4 \to \mathbb{C}^m$ such that all distribution derivatives of order less than or equal to s are functions in H_τ. We have $H_{\tau,0} = H_\tau$ and $H_{\tau,s} \subset H_{\tau,s'}$ for $s > s'$. $H_{0,s}$ are the usual Sobolev spaces. An inner product for $H_{\tau,s}$ may be defined either by

$$(f,g)_{\tau,s} = \sum_{|\alpha| \leqslant s} (\partial^\alpha f, \partial^\alpha g)_\tau, \tag{2.1}$$

where $\alpha = (\alpha_0,\ldots,\alpha_3)$ is a multi-index, $|\alpha| = \alpha_0 + \cdots + \alpha_3$, and $\partial^\alpha = \partial_0^{\alpha_0} \cdots \partial_3^{\alpha_3}$; or else inductively by

$$(f,g)_{\tau,s+1} = (f,g)_{\tau,s} + \sum_{\mu=0}^{3} (\partial_\mu f, \partial_\mu g)_{\tau,s}. \tag{2.2}$$

The associated norms are equivalent, and we generally make the second choice. That $H_{\tau,s}$ is complete, and hence a Hilbert space follows by the usual proof for $\tau = 0$. It is also straightforward to show that $C_0^\infty(R^4, \mathbb{C}^m)$ is dense in $H_{\tau,s}$.

The inner product in $H_{\tau,s}$ can also be expressed in terms of the Fourier transform. For example, if $f,g \in C_0^\infty$ the Fourier transforms $\hat{f},\hat{g}$ are entire functions and by Parseval's theorem we obtain for $s \geqslant 0$

$$(f,g)_{\tau,s} = \int \sum_j \overline{\hat{f}_j(p - i\tau n)} \hat{g}_j(p - i\tau n)(1 + |p - i\tau n|^2)^s dp.$$

This integral also makes sense for $s < 0$ and we may define $H_{\tau,s}$ for $s < 0$ as the completion of C_0^∞ with respect to the associated norm. We further note that by shifting the contour of integration we have

$$\langle f,g\rangle = \int \sum_j \hat{f}_j(p - i\tau n) \hat{g}_j(-p + i\tau n) dp.$$

This gives the estimate $|\langle f,g\rangle| \leqslant \|f\|_{\tau,s} \|g\|_{-\tau,-s}$ for any τ,s. Hence the form $f,g \to \langle f,g\rangle$ extends to $H_{\tau,s} \times H_{-\tau,-s}$ with the same bound. It is also true that $(H_{\tau,s})' = H_{-\tau,-s}$ with this pairing, but this fact will not be needed.

It is straightforward to show that a matrix of linear differential operators of order m defines a continuous operator from $H_{\tau,s+m}$ to $H_{\tau,s}$ $(s = 0,1,2,\cdots)$ provided that the derivatives of the coefficients are bounded. Thus, if $A_\mu \in C_0^\infty(R^4, \mathscr{L}(\mathbb{C}^m))$ [$\mathscr{L}(\mathbb{C}^m) = m \times m$ matrices], we have that $\Box_A + m^2$ is continuous from $H_{\tau,s+2}$ to $H_{\tau,s}$. The following lemma shows that this operator has an inverse.

Lemma 2.1: Let $|\tau|$ be sufficiently large. Then there exist operators denoted E_A^+ for $\tau > 0$ or E_A^- for $\tau < 0$ which map $H_{\tau,s}$ continuously to $H_{\tau,s+1}$, $s = 0,1,2,\cdots$, and satisfy

$$(\Box_A + m^2) E_A^\pm = E_A^\pm (\Box_A + m^2) = I.$$

Proof: First consider the case $A = 0$. It is well known that the Klein–Gordon operator $\Box_0 + m^2$ has fundamental solutions $E_0^\pm : C_0^\infty \to C^\infty$. In fact, they are given explicitly by taking $\pm\tau < 0$ and defining

$$(E_0^\pm f)(x) = (2\pi)^{-2} \int e^{i(p - i\tau n)x} \hat{f}(p - i\tau n) \times [-(p - i\tau n)_\mu (p - i\tau n)^\mu + m^2]^{-1} dp. \tag{2.3}$$

Corresponding to the fact that $\Box_0 + m^2$ is strictly hyperbolic, the denominator never vanishes for $\tau \neq 0$. and moreover, we have the basic *a priori* estimate for $|\tau|$ sufficiently large:

$$|-(p - i\tau n)_\mu (p - i\tau n)^\mu + m^2|^2 \geqslant C\tau^2 (1 + |p - i\tau n|^2).$$

(See, for example, Treves, [11] Sec. II.5.) It follows that

$$\|E_0^{\pm} f\|_{\tau,s+1}^2$$
$$= \int |\hat{f}(p - i\tau n)|^2 |-(p - i\tau n)_\mu (p - i\tau n)^\mu + m^2|^{-2} \times (1 + |p - i\tau n|^2)^{s+1} dp$$
$$\leqslant C^{-1}\tau^{-2} \int |\hat{f}(p - i\tau n)|^2 (1 + |p - i\tau n|^2)^s dp$$
$$= C^{-1}\tau^{-2} \|f\|_{\tau,s}^2.$$

Thus $E_0^{\pm}$ defines a continuous operator from $H_{\tau,s}$ to $H_{\tau,s+1}$ with a norm which is $\mathcal{O}(|\tau|^{-1})$ as $|\tau| \to \infty$.

We now write

$$\Box_A + m^2 = \Box_0 + m^2 + \mathscr{A},$$

where

$$\mathscr{A} = 2iA_\mu \partial^\mu + i(\partial_\mu A^\mu)(x) - A_\mu(x) A^\mu(x). \tag{2.4}$$

The operator $\mathscr{A}$ maps $H_{\tau,s+1}$ to $H_{\tau,s}$ with a norm which is bounded uniformly in τ. Thus $\mathscr{A} E_0^{\pm}$ is a bounded operator on $H_{\tau,s}$ with a norm which is $\mathcal{O}(|\tau|^{-1})$ as $|\tau| \to \infty$. For $|\tau|$ large enough $\|\mathscr{A} E_0^{\pm}\| < 1$, and we may define $E_A^{\pm}$ by the convergent power series

$$E_A^{\pm} = E_0^{\pm} \left(\sum_{k=0}^{\infty} (-1)^k (\mathscr{A} E_0^{\pm})^k \right). \tag{2.5}$$

Then $E_A^{\pm}$ satisfies the conditions of the lemma. ■

Lemma 2.2: Let $H_{\tau,\infty} = \cap_s H_{\tau,s}$ with topology defined by the norms $\|\cdot\|_{\tau,s}$. Then we have the topological inclusions

$$\mathscr{D} \subset H_{\tau,\infty} \subset \mathscr{E}$$

where $\mathscr{D} = C_0^\infty(R^4, \mathbb{C}^m)$ and $\mathscr{E} = C^\infty(R^4, \mathbb{C}^m)$ with the usual topologies.

Proof: $\mathscr{D} \subset H_{\tau,\infty}$ is clear. We must show that the injection is continuous. It suffices to show that the restriction to $\mathscr{D}_K = C_0^\infty(K, \mathbb{C}^m)$ is continuous for compact K. But this follows from the fact that, for any s, there exist constants C_1, C_2 such that for $f \in \mathscr{D}_K$

$$\|f\|_{\tau,s} \leqslant C_1 \|f\|_{0,s} \leqslant C_2 \sum_{|\alpha| \leqslant s} \sup_{x \in K} |(\partial^\alpha f)(x)|.$$

To see that $H_{\tau,\infty} \subset \mathscr{E}$, note that $f \in H_{\tau,\infty}$ implies $\phi f \in H_{0,\infty}$ for any $\phi \in C_0^\infty(R^4)$. By Sobolev's lemma we have $\phi f \in \mathscr{E}$ and hence $f \in \mathscr{E}$. The injection map from $H_{\tau,\infty}$ to $\mathscr{E}$ is continuous since for any compact K and $s \geqslant 0$ there exist C_1, C_2, s' such that for $f \in H_{\tau,\infty}$

$$\sum_{|\alpha| \leqslant s} \sup_{x \in K} |(\partial^\alpha f)(x)| \leqslant C_1 \|\phi f\|_{0,s'} \leqslant C_2 \|\phi f\|_{\tau,s'} \leqslant C_2 \|f\|_{\tau,s'},$$

where $\phi \equiv 1$ on K. ∎

Theorem 2.3: Let $A_\mu \in C_0^\infty(R^4, \mathscr{L}(\mathbb{C}^m))$. Then

(1) There exist operators $E_A^{\pm}$ mapping $\mathscr{D}$ continuously to $\mathscr{E}$ such that

(a) $(\Box_A + m^2) E_A^{\pm} = E_A^{\pm} (\Box_A + m^2) = I$,

(b) Supp$(E_A^{\pm} f)$ is contained in he future (past) shadow of supp$f$.

These are the unique operators from $\mathscr{D}$ to $\mathscr{D}'$ with these properties.

(2) $E_A^{\pm}$ extends to a continuous operator from $\mathscr{E}'$ to $\mathscr{D}'$ satisfying (a), (b), and

$$E_A^{\pm} = (E_{-A'}^{\mp})'.$$

Proof: For existence pick any $|\tau| > 0$ sufficiently large. Then Lemma 2.1 gives $E_A^{\pm}$ as a continuous operator on $H_{\tau,\infty}$ satisfying (a). Property (b) is also satisfied by (2.5) since $E_0^{\pm}$ satisfies it and $\mathscr{A}$ shrinks supports. By Lemma 2.2 the restriction of $E_A^{\pm}$ is continuous from $\mathscr{D}$ to $\mathscr{E}$. Properties (a), (b) are still satisfied.

For uniqueness we first claim that if $u \in \mathscr{D}'$, $(\Box_A + m^2)u = 0$, and $u = 0$ in the distant past (i.e., x_0 sufficiently negative), then $u = 0$. This follows from $(\Box_A + m^2)' = (\Box_{-A'} + m^2)$, for we have any $f \in \mathscr{D}$

$$\langle u, f \rangle = \langle u, (\Box_{-A'} + m^2) E_{-A'}^{-} f \rangle$$
$$= \langle (\Box_A + m^2) u, E_{-A'}^{-} f \rangle$$
$$= 0.$$

Note that $E_{-A'}^{-} f$ is a proper test function for u since it has support in the past of supp$f$. Now, if $E_A^{+}, \tilde{E}_A^{+}$ are two fundamental solutions, we set $u = (E_A^{+} - \tilde{E}_A^{+}) f, f \in \mathscr{D}$, and conclude $u = 0$. Thus E_A^{+} is unique.

For (2) it suffices to show that $(E_{-A'}^{-})' : \mathscr{E}' \to \mathscr{D}'$ satisfies (a) and (b), for then it agrees with E_A^{+} on $\mathscr{D}$ and provides the extension. But (a) follows by the adjoint of (a) for $-A'$, and (b) follows since if $f \in \mathscr{E}'$, $g \in \mathscr{D}$, and supp$g$ does not intersect the future of supp$f$, then supp$f$ does not intersect the past of supp$g$ and so

$$\langle g, (E_{-A'}^{-})' f \rangle = \langle E_{-A'}^{-} g, f \rangle = 0.$$ ■

Remarks: (1) From the definition we have the identities on $H_{\tau,\infty}$:

$$E_A^{\pm} = E_A^{\pm} - E_A^{\pm} \mathscr{A} E_0^{\pm} = E_0^{\pm} - E_0^{\pm} \mathscr{A} E_A^{\pm}. \tag{2.6}$$

Since $\mathscr{A}$ is continuous from $\mathscr{E}$ to $\mathscr{D}$, the identity also makes sense and holds in $\mathscr{L}(\mathscr{D}, \mathscr{E})$. Also $\mathscr{A} : \mathscr{D}' \to \mathscr{E}'$, and by taking adjoints we have the identity in $\mathscr{L}(\mathscr{E}', \mathscr{D}')$.

(2) Suppose that u_{in} or u_{out} in $\mathscr{D}'$ is a solution of the free equation $(\Box_0 + m^2)u = 0$. Then we can construct a solution of $(\Box_A + m^2)u = 0$, which agrees with u_{in} in the past or u_{out} in the future by

$$u = (1 - E_A^{+} \mathscr{A}) u_{\text{in}}, \quad u = (1 - E_A^{-} \mathscr{A}) u_{\text{out}}. \tag{2.7}$$

These are unique by the proof of Theorem 2.3. On the other hand, given a solution u, we can construct free solutions u_{in} or u_{out} which agree in the past or future by

$$u_{\text{in}} = (1 + E_0^{+} \mathscr{A}) u, \quad u_{\text{out}} = (1 + E_0^{-} A) u. \tag{2.8}$$

(3) The preceding results will be sufficient for the construction of the quantum field theory. We now turn to the estimate on $E_A^{\pm}$ which is needed for a unitary S operator.

Theorem 2.4: Let $f_1, f_2 \in \mathscr{D}$ and define $\exp_p(x) = e^{-ipx}$. Then for any $s \geqslant 0$ there is a constant K such that

$$|\langle \exp_p f_1, E_A^{\pm} \exp_q f_2 \rangle| \leqslant K (1 + |p + q|^2)^{-s/2}.$$

Proof: Let $E = E_A^{\pm}$ and let $E_q = \exp_{-q} E \exp_q$, regarded as an operator from $H_{\tau,s}$ to itself. Then

$$|\langle \exp_p f_1, E \exp_q f_2 \rangle| = |\langle \exp_{p+q} f_1, E_q f_2 \rangle|$$
$$\leqslant \|\exp_{p+q} f_1\|_{-\tau,-s} \|E_q\|_{\tau,s} \|f_2\|_{\tau,s}$$

The first factor is $\mathcal{O}(1+|p+q|^2)^{-s/2}$, for we have the general estimate

$$\begin{aligned}\|\exp_p f\|^2_{-\tau,-s} &= \int |\hat{f}(p+p'+i\tau n)|^2(1+|p'+i\tau n|^2)^{-s}dp' \\ &\leqslant K'\int(1+|p+p'|^2)^{-s}(1+|p'|^2)^{-s}dp' \\ &\leqslant K(1+|p|^2)^{-s}.\end{aligned}$$

Thus we have reduced the problem to showing that $\|E_q\|_{\tau,s}$ is bounded in q. For $s=0$ this is trivial, and the idea is to relate the bound for general s to bounds at $s=0$. To accomplish this, we introduce the multicommutator

$$E_{q;\mu_1,\ldots,\mu_n} = [\partial_{\mu_n},[\ldots,[\partial_{\mu_1},E_q]\cdots]],$$

regarded initially as an operator on $H_{\tau,\infty}$. We claim that for any s, n, and $(\mu_1,\ldots,\mu_n)$, $E_{q;\mu_1,\ldots,\mu_n}$ extends to a bounded operator on $H_{\tau,s}$ with norm bounded uniformly in q. The result we need is the special case $n=0$.

First consider $E_{\mu_1,\ldots,\mu_n} = E_{q=0;\mu_1,\ldots,\mu_n}$. Since $[\partial_\mu, \square_A + m^2] = \partial_\mu \mathscr{A}$, we have $[\partial_\mu, E] = -E(\partial_\mu \mathscr{A})E$. Thus we can write $E_{\mu_1,\ldots,\mu_n}$ as a sum of terms of the form

$$E(\partial^{\alpha_1})\mathscr{A}E\cdots(\partial^{\alpha_r})\mathscr{A})E,$$

where α_i is a multi-index. Since $E: H_{\tau,0}\to H_{\tau,1}$ and $\partial^\alpha \mathscr{A}: H_{\tau,1}\to H_{\tau,0}$, these expressions are bounded operators on $H_{\tau,0}$ and hence so is $E_{\mu_1,\ldots,\mu_n}$. For any operator $\mathcal{O}$, if $\mathcal{O}_q = \exp_{-q}\mathcal{O}\exp_q$, we have $[\partial_\mu, \mathcal{O}_q] = [\partial_\mu, \mathcal{O}]_q$. Thus $E_{q;\mu_1,\ldots,\mu_n} = (E_{\mu_1,\ldots,\mu_n})_q$ is also a bounded operator on $H_{\tau,0}$ and

$$\|E_{q;\mu_1,\ldots,\mu_n}\|_{\tau,0} = \|(E_{\mu_1,\ldots,\mu_n})_q\|_{\tau,0} = \|E_{\mu_1,\ldots,\mu_n}\|_{\tau,0}$$

is independent of q. Thus the claim is proved for $s=0$.

The claim for general s is proved by induction. Suppose it is true for s. Then it is true for $s+1$ by the estimate for $f\in\mathscr{H}_{\tau,\infty}$

$$\begin{aligned}&\|E_{q;\mu_1,\ldots,\mu_n}f\|^2_{\tau,s+1} \\ &= \|E_{q;\mu_1,\ldots,\mu_n}f\|^2_{\tau,s} + \sum_\mu \|\partial E_{\mu q;\mu_1,\ldots,\mu_n}f\|^2_{\tau,s} \\ &\leqslant 2\|E_{q;\mu_1,\ldots,\mu_n}\|^2_{\tau,s}\|f\|^2_{\tau,s+1} \\ &\quad + 2\sum_\mu \|E_{q;\mu_1,\ldots,\mu_n,\mu}\|^2_{\tau,s}\|f\|^2_{\tau,s} \\ &\leqslant \mathcal{O}(1)\|f\|^2_{\tau,s+1}. \quad \blacksquare\end{aligned}$$

III. THE QUANTUM PROBLEM

A. We now turn to the quantum problem, mainly following the general formalism of Wightman.[3,4] The first step is to review the definition of the free charged scalar field with distinct particles and antiparticles. The one-particle Hilbert space is defined to be $\mathscr{H}_1 = L_2(R^3,\mathbb{C}^m,d\mu(\mathbf{p}))$, where $d\mu(\mathbf{p}) = [2\omega(\mathbf{p})]^{-1}d\mathbf{p}$ and $\omega(\mathbf{p}) = (|\mathbf{p}|^2+m^2)^{1/2}$. The n-particle space is the symmetric tensor product $\mathscr{H}_n = \otimes_s^n \mathscr{H}_1$ and we set $\mathscr{H} = \oplus_{n=0}^\infty \mathscr{H}_n$. The full Hilbert space $\mathscr{F}$ is a tensor product of particle and antiparticle spaces:

$$\mathscr{F} = \mathscr{H}\otimes\mathscr{H} = \bigoplus_{n,m=0}^\infty \mathscr{H}_n\otimes\mathscr{H}_m.$$

For $h\in\mathscr{H}_1$, let $\alpha(h)$ and $\alpha^*(h)$ be the usual annihilation and creation operators defined on the finite particle vectors in $\mathscr{H}$, and then define particle and antiparticle operators on $\mathscr{F}$ by

$$a(h) = \alpha(h)\otimes I, \quad a^*(h) = \alpha^*(h)\otimes I,$$
$$b(h) = I\otimes\alpha(h), \quad b^*(h) = I\otimes\alpha^*(h).$$

Then

$$[a(h)]^* = a^*(\bar{h}), \quad [b(h)]^* = b^*(\bar{h}), \tag{3.1}$$

and we have the canonical commutation relations (CCR),

$$\begin{aligned}[a(h),a^*(h')] &= [b(h),b^*(h')] \\ &= \int \sum_j h_j(\mathbf{p})h'_j(\mathbf{p})\,d\mu(\mathbf{p}).\end{aligned} \tag{3.2}$$

Furthermore, there is a unique vacuum Ω_0 (the no-particle state) satisfying

$$a(h)\Omega_0 = b(h)\Omega_0 = 0. \tag{3.3}$$

Now define $\Pi^\pm: \mathscr{D}\to\mathscr{H}_1$ by Fourier transformation (now with Lorentz inner product) followed by restriction to the mass shell:

$$(\Pi^\pm f)(\mathbf{p}) = (2\pi)^{1/2}\hat{f}(\pm\omega(\mathbf{p}),\pm\mathbf{p}).$$

The free field operator is defined by

$$\begin{aligned}\phi_0(f) &= a(\Pi^+ f) + b^*(C\Pi^- f), \\ \phi_0^*(f) &= a^*(\Pi^- f) + b(\bar{C}\Pi^+ f).\end{aligned} \tag{3.4}$$

Here we allow for a charge conjugation operator C on $\mathbb{C}^m$, which is unitary and satisfies $C^2 = I$. Then we have

$$[\phi_0(f)]^* = \phi_0^*(\bar{f}), \tag{3.5}$$

the free field equation

$$\phi_0((\square_0 + m^2)f) = 0, \tag{3.6}$$

and the commutation relations

$$[\phi_0(f),\phi_0^*(g)] = (1/i)\langle f, E_0 g\rangle, \tag{3.7}$$

where $E_0 \equiv E_0^+ - E_0^-$ is identified with the usual commutator function $\Delta = \Delta_+ + \Delta_-$.

Hereafter, if $T:\mathscr{D}\to\mathscr{D}$ has adjoint $T':\mathscr{D}'\to\mathscr{D}'$, we write $T'\phi$ to mean the operator valued distribution $(T'\phi)(f) = \phi(Tf)$. Thus (3.6), for example, is written $(\square_0 + m^2)\phi = 0$.

B. Now let ϕ_{in} be a free field operator as defined above. We want to define a field operator ϕ solving (1.1) with ϕ_{in} as data in the distant past. According to the discussion following Theorem 2.3, we therefore define for $f\in\mathscr{D}$

$$\phi(f) = [(1 - E_A^+\mathscr{A})\phi_{\text{in}}](f) = \phi_{\text{in}}[(1 - \mathscr{A}'E_{-A'}^-)f],$$

where $\mathscr{A}$ is defined by (2.4).

We also define

$$\phi^*(f) = [\phi(\bar{f})]^* = \phi_{\text{in}}^*[(1 - \mathscr{A}E_A^-)f].$$

Here we use $(1 - E_A^+\mathscr{A})^* = 1 - \mathscr{A}E_A^-$ (T^* means $\overline{T'}$), which follows from $A_\mu^* = A_\mu$.

Theorem 3.1: Let $A_\mu\in C_0^\infty(R^4,\rho(G'))$. Then

(a) (Existence) $\phi(f)$ is a continuous function of $f\in\mathscr{D}$ satisfying $(\square_A + m^2)\phi = 0$ and $\phi = \phi_{\text{in}}$ in the distant past.

(b) (Locality) $[\phi(f),\phi^*(g)] = (1/i)\langle f, E_A g\rangle$, where

$E_A = E_A^+ - E_A^-$. In particular, spacelike separated fields commute.

Remark: Continuous means continuous in whatever sense the free field is continuous. For example, for finite particle vectors $\psi,\chi\in\mathcal{F}$, the function $f\to(\psi,\phi(f)\chi)$ is continuous.

Proof: Part (a) follows directly from the definitions. Given the free field commutation relations, part (b) follows once we show

$$(1 - E_A^+ \mathcal{A})E_0(1 - \mathcal{A}E_A^-) = E_A. \tag{3.8}$$

However, using (2.6), we have

$$(1 - E_A^+ \mathcal{A})E_0^{\pm}(1 - \mathcal{A}E_A^-) = E_A^{\pm} - E_A^+ \mathcal{A}E_A^-.$$

Taking the difference of the (+) equation and the (−) equation yields (3.8), completing the proof. □

C. According to the remarks in Sec. II, we now define for $f\in\mathcal{D}$

$$\phi_{\text{out}}(f) = [(1 + E_0^- \mathcal{A})\phi](f), \quad \phi_{\text{out}}^*(f) = [\phi_{\text{out}}(\bar{f})]^* \tag{3.9}$$

Then ϕ_{out} satisfies the free field equation (3.6), and $\phi = \phi_{\text{out}}$ in the distant future. Furthermore, ϕ_{out} satisfies the free field commutation relations (3.7) by an argument similar to that of Theorem 3.1.

The in and out fields are related by

$$\phi_{\text{out}}(f) = (\mathcal{R}\phi_{\text{in}})(f), \tag{3.10}$$

where $\mathcal{R}:\mathcal{D}'\to\mathcal{D}'$ is given by

$$\begin{aligned}\mathcal{R} &= (1 + E_0^- \mathcal{A})(1 - E_A^+ \mathcal{A}) \\ &= 1 - E_0(\mathcal{A} - \mathcal{A}E_A^+ \mathcal{A}) \\ &\equiv 1 - E_0 J.\end{aligned} \tag{3.11}$$

Here the first step follows by replacing E_0^- by $E_0^+ - E_0$ and using (2.6), and the second step is the definition of J. At this point we note that $\phi_{\text{out}}(f) = \phi_{\text{in}}(\mathcal{R}'f)$ allows the definition of ϕ_{out} to be continuously extended to the Schwartz space $\mathcal{S} = \mathcal{S}(R^4,\mathbb{C}^m)$. This follows since E_0', J', and hence $\mathcal{R}'$ are continuous operators on $\mathcal{S}$, and ϕ_{in} is defined and continuous in $\mathcal{S}$.

Now we may introduce creation and annihilation operators. For any h in $\mathcal{S}(R^3,\mathbb{C}^m)$ there exists h^+ in $\mathcal{S}(R^4, \mathbb{C}_m)$ such that $\Pi^+h^+ = h$ and $\Pi^-h^+ = 0$. Then we define

$$a_{\text{out}}(h) = \phi_{\text{out}}(h^+), \quad b_{\text{out}}(h) = \phi_{\text{out}}^*(Ch^+)$$

and $a_{\text{out}}^*, b_{\text{out}}^*$ so that the adjoint relation (3.1) is satisfied. These definitions are independent of the choice of h^+. These operators satisfy the CCR (3.2), and ϕ_{out} and $a_{\text{out}}, b_{\text{out}}$ are related by (3.4). Now all that is lacking for ϕ_{out} to be a standard free field is a vacuum vector. This question is resolved by the next theorem.

Let us note that (3.10) becomes

$$\begin{aligned}a_{\text{out}}(h) &= a_{\text{in}}(\Pi^+\mathcal{R}'h^+) + b_{\text{in}}^*(C\Pi^-\mathcal{R}'h^+), \\ b_{\text{out}}(h) &= b_{\text{in}}(\bar{C}\Pi^+\mathcal{R}^*Ch^+) + a_{\text{in}}^*(\Pi^-\mathcal{R}^*Ch^+).\end{aligned} \tag{3.12}$$

We now show that this transformation is unitarily implementable.

Theorem 3.2: These exists a unitary operator S on $\mathcal{F}$ such that $a_{\text{out}} = S^{-1}a_{\text{in}}S$, $b_{\text{out}} = S^{-1}b_{\text{in}}S$ and hence $\phi_{\text{out}} = S^{-1}\phi_{\text{in}}S$.

Proof: According to a standard theorem a CCR preserving transformation of the general form (3.12) is unitarily implementable if the off-diagonal operators $h\to C\Pi^-\mathcal{R}'h^+$ and $h\to\Pi^-\mathcal{R}^*Ch^+$ are Hilbert–Schmidt on $\mathcal{H}_1$.

To see this first consider $J = \mathcal{A} - \mathcal{A}E_A^+\mathcal{A}$. The distribution kernel $J(x,y)$ has compact support and thus has a Fourier transform $\hat{J}(p,q)$ which is matrix of smooth functions. Typical terms in $\hat{J}_{ij}(p,q)$ are (up to a constant)

$$(\hat{A}_\mu)_{ij}(p+q)q^\mu, \quad p^\mu\langle\exp_p(A_\mu)_{i(\cdot)}, E_A^+ \exp_q(A_\nu)_{(\cdot)j}\rangle q^\nu.$$

Here $(A_\mu)_{i(\cdot)}$ denotes the ith row of A_μ, etc. By Theorem 2.4 we have the estimate for any s

$$|\hat{J}_{ij}(p,q)| \leqslant K_s(1 + |p|^2)^{1/2}(1 + |q|^2)^{1/2}(1 + |p+q|^2)^{-s/2}, \tag{3.13}$$

Now using $\mathcal{R}' = 1 + J'E_0$ and

$$\langle f, E_0h^+\rangle = (2\pi)^{1/2}i\int\sum_j \hat{f}_j(-\omega(\mathbf{q}), -\mathbf{q})h_j(\mathbf{q})d\mu(\mathbf{q})$$

and $(J')\hat{}_{ij}(p,q) = \hat{J}_{ji}(q,p)$, we have

$$\begin{aligned}&(\Pi^-\mathcal{R}'h^+)_i(\mathbf{p}) \\ &\quad = 2\pi i\int\sum_j \hat{J}_{ji}(-\omega(\mathbf{q}), -\mathbf{p}; -\omega(\mathbf{p}), -\mathbf{p})h_j(\mathbf{q})d\mu(\mathbf{q}).\end{aligned}$$

By (3.13) the kernel is rapidly decreasing in $|\mathbf{p}|$, $|\mathbf{q}|$ and so the operator $h\to C\Pi^-\mathcal{R}'h^+$ is Hilbert–Schmidt. Similarly the other operator is Hilbert–Schmidt. □

Discussion: If Ω_{in} is the vacuum for $a_{\text{in}}, b_{\text{in}}$, then $\Omega_{\text{out}} = S^{-1}\Omega_{\text{in}}$ provides a unique vacuum for $a_{\text{out}}, b_{\text{out}}$. Now states of the form

$$\Phi_{\text{out}} = \prod_{i=1}^{s} a_{\text{out}}^*(h_i)\prod_{j=1}^{r} b_{\text{out}}^*(g_j)\Omega_{\text{out}} \tag{3.14}$$

may be interpreted as containing s particles and r antiparticles in the distant future, just as a similar interpretation is held for

$$\Psi_{\text{in}} = \prod_{i=1}^{s'} a_{\text{in}}^*(h_i')\prod_{j=1}^{r'} b_{\text{in}}^*(g_j')\Omega_{\text{in}} \tag{3.15}$$

in the distant past. Scattering is described by the amplitudes

$$(\Phi_{\text{out}}, \Psi_{\text{in}}) = (\Phi_{\text{in}}, S\Psi_{\text{in}}). \tag{3.16}$$

The qualitative features of scattering are discussed in Wightman.[4]

D. Finally, we consider the behavior of our theory under gauge transformations. We let $C_0^\infty(R^4,\rho(G))$ be the space of C^∞ functions $\gamma:R^4\to\rho(G)$ such that $\gamma =$ identity off a compact set.

Theorem 3.3: Let $A_\mu, A_\mu'\in C_0^\infty(R^4,\rho(G'))$ be gauge potentials with associated field operators $\phi(f), \phi'(f)$ and scattering operators S, S'. If $(iA_\mu') = \gamma^{-1}(iA_\mu)\gamma + \gamma^{-1}\partial_\mu\gamma$ for some $\gamma\in C_0^\infty(R^4,\rho(G))$, then $\phi'(f) = (\gamma^{-1}\phi)(f)$ and $S' = S$.

Proof: We must show that $\phi - \gamma\phi' = 0$. This operator is defined from ϕ_{in} by

$$\begin{aligned}&(1 - E_A^+\mathcal{A}) - \gamma(1 - E_{A'}^+\mathcal{A}') \\ &\quad = (E_A^+ - \gamma E_{A'}^+)(\Box_0 + m^2) \\ &\quad = E_A^+(1 - \gamma)(\Box_0 + m^2).\end{aligned}$$

Here we use (2.6) and also the identity $E_{A'}^{\pm} = \gamma^{-1}E_A^{\pm}\gamma$

which follows from the same property for $\Box_A + m^2$ and Theorem 2.3. Thus

$$\phi(f) - (\gamma\phi')(f) = [(\Box_0 + m^2)\phi_{\text{in}}][(1-\gamma')(E_A^+)'f] = 0$$

since $(1-\gamma')(E_A^+)'f$ is in $\mathscr{D}$.

Next we claim $\phi_{\text{out}} - \phi'_{\text{out}} = 0$. This operator is defined from $\phi = \gamma\phi'$ by

$$\begin{aligned} &= (1 + E_0^- \mathscr{A}) - (1 + E_0^- \mathscr{A}')\gamma^{-1} \\ &= E_0^-[(\Box_A + m^2) - (\Box_{A'} + m^2)\gamma^{-1}] \\ &= E_0^-(1-\gamma^{-1})(\Box_A + m^2). \end{aligned}$$

Thus

$$\phi_{\text{out}}(f) - \phi'_{\text{out}}(f) = [(\Box_A + m^2)\phi][(1-\bar{\gamma})E_0^+ f] = 0$$

since $(1-\bar{\gamma})E_0^+ f$ is in $\mathscr{D}$.

Finally $\phi_{\text{out}} = \phi'_{\text{out}}$ implies $S = S'$ on sums of states of the form (3.15). Since this is a dense set we conclude $S = S'$.

Discussion: To interpret this theorem, we return to the fibre bundle language. Now, however, only gauges in $C_0^\infty(R^4, G)$ are admitted. We suppose that $\omega = \{A_g^\mu\}$ is a connection which is trivial off a compact set in the sense that for some (and hence all) $g \in C_0^\infty(R^4, G)$ the field A_g^μ is in $C_0^\infty(R^4, \rho(G'))$. Let $\phi_g(f)$ be the corresponding family of field operators solving $(\Box_{A_g} + m^2)\phi_g = 0$.

The theorem says that $\phi_{g'}(f) = (\gamma^{-1}\phi_g)(f)$ just as in (1.2); i.e., $\phi_g(f)$ has the structure of a cross section of E. We can make this more precise by noting that cross sections $H = \{h_g\}$ of E can be identified with antilinear functionals on cross sections $F = \{f_g\}$ with compact support [denoted $C_0^\infty(E)$] by $H(F) = \langle h_g \bar{f}_g \rangle$. The integral is independent of g since $\rho(G)$ is unitary. Correspondingly, we can define a field operator Φ as an operator valued function on $C_0^\infty(E)$ by $\Phi(F) = \phi_g(\bar{f}_g)$ which is also independent of g. It is Φ which is the basic dynamical entity. Corresponding to the identity $(\mathscr{D}_\omega H)(F) = H(\mathscr{D}_\omega F)$, we define $(\mathscr{D}_\omega \Phi)(F) = \phi(\mathscr{D}_\omega F)$ for the field operator. Then it is easy to check that Φ satisfies the dynamical equation $\mathscr{D}_\omega \Phi = 0$.

The theorem also says that the S operator S_g for A_g^μ is independent of g and so depends only on ω. Since ω is the basic model for the external field and S is the basic observable, the attainment of this result has been one of our main goals.

Note added in manuscript: After the completion of this manuscript, a conversation with A. Wightman made me aware of the proceedings of the 1977 Erice Conference which have recently appeared (G. Velo and A. Wightman, Eds., *Invariant Wave Equations*, Lecture Notes in Physics, Vol. 73 (Springer-Verlag, Berlin-Heidelberg-New York, 1978). These lectures touch on many of the same areas as the present paper and are recommended to the reader. In particular Ruijsenaars announces results similar to ours for the electromagnetic case. Also Seiler's lectures [5] appear in this collection, and Wightman's lectures contain a proof of the fact that the Hilbert–Schmidt property suffices for the existence of S as quoted in Theorem 3.2.

[1] A. Trautman, "Fibre bundles associated with spacetime," Rep. Math. Phys. **1**, 29–62 (1970).
[2] W. Drechsler and M.E. Mayer, *Fiber Bundle Techniques in Gauge Theories* (Springer, New York, 1977).
[3] A. Wightman, "Partial Differential Equations and Relativistic Quantum Field Theory, in *Lecture Series in Differential Equations II*, edited by A.K. Aziz (Van Nostrand, New York, 1969).
[4] A. Wightman, "Relativistic Wave Equations as Singular Hyperbolic Systems," in Proceedings of the Symposium *in Pure Mathematics* (AMS, Providence, 1973), Vol. 23.
[5] R. Seiler, "Particles with spin $S \leqslant 1$ in an External Field," Berlin preprint, 1977.
[6] B. Schroer, R. Seiler, and J. Swieca, "Problems of Stability for Quantum fields in External Time-dependent Potentials," Phys. Rev. D **2**, 2927 (1970).
[7] J. Belissard, "Quantized Fields in Interaction with External Fields," Comm. Math. Phys. **41**, 235–266 (1975) and **46**, 53–74 (1976).
[8] R. Schrader, Towards a Constructive Approach of a Gauge Invariant Massive $\mathscr{P}(\phi)_2$ Theory," Commun. Math. Phys. **58**, 299–312 (1978).
[9] J. Dimock, "Scalar quantum field in an external gravitational field," preprint.
[10] J. Leray, Hyperbolic Differential Equations, Princeton Lecture Notes, 1953.
[11] F. Treves, *Linear Partial Differential Equations* (Gordon and Breach, New York, 1970).

On a common property of gravitational and chiral exponential Green functions

H. E. De Meyer[a)]

Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 271, Gent, Belgium
(Received 24 January 1979)

The dependence of gravitational and chiral exponential Green functions on a "gauge" parameter which enters the free propagation function of the basic matrix field, is analyzed. It is proven that by an appropriate choice of this parameter, the Green functions can be made ambiguity-free.

1. INTRODUCTION

Nonlinear chiral theories and gravity modified field theories have the common property of containing, in the Lagrangian density, nonpolynomial functions of a matrix field. By using localizable exponential parametrization, the problem of calculating vacuum expectation values of time ordered products of these functions can then in its simplest form be mathematically formulated as follows.

Given a ν-dimensional real symmetrical (graviton) matrix field $\phi_{\alpha\beta}(x)$, $(\alpha,\beta = 1,2,...,\nu)$, which propagates as

$$\langle\phi_{\alpha\beta}(x),\phi_{\gamma\delta}(y)\rangle = \tfrac{1}{2}(\delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma} - 2c\delta_{\alpha\beta}\delta_{\gamma\delta})\Delta(x-y), \quad (1.1)$$

and a ν-dimensional Hermitian (chiral) matrix field $\bar{\phi}_{\alpha\beta}(x)$, $(\alpha,\beta = 1,2,...,\nu)$, which propagates as

$$\langle\bar{\phi}_{\alpha\beta}(x),\bar{\phi}_{\gamma\delta}(y)\rangle = (\delta_{\alpha\gamma}\delta_{\beta\delta} - \bar{c}\delta_{\alpha\beta}\delta_{\gamma\delta})\Delta(x-y), \quad (1.2)$$

find, by taking into account Wick's reduction theorem, closed expressions for the traced Green functions

$$T^{(\nu)}_{k_1,k_2,...,k_n}[c,\{\Delta_{ij} \mid i,j\in\mathscr{J}_n; i<j\}] = \langle[\mathrm{Tr}\exp(\kappa\phi(x_1))]^{k_1},[\mathrm{Tr}\exp(\kappa\phi(x_2))]^{k_2},..., [\mathrm{Tr}\exp(\kappa\phi(x_n))]^{k_n}\rangle, \quad (1.3)$$

and

$$\bar{T}^{(\nu)}_{k_1,k_2,...,k_n}[\bar{c},\{\Delta_{ij} \mid i,j\in\mathscr{J}_n; i<j\}] = \langle[\mathrm{Tr}\exp(\kappa\bar{\phi}(x_1))]^{k_1},[\mathrm{Tr}\exp(\kappa\bar{\phi}(x_2))]^{k_2},..., [\mathrm{Tr}\exp(\kappa\bar{\phi}(x_n))]^{k_n}\rangle, \quad (1.4)$$

$\mathscr{J}_n$ denoting the set formed by the first n positive integers. $\Delta_{ij} \equiv \Delta(x_i - x_y)$ stands for a propagator of the real scalar massless Klein–Gordon field, κ is an arbitrary constant, and the fields are as usual assumed to be defined over four-dimensional Euclidean space–time. In (1.1) the real parameter c is the so-called gauge parameter, related to the weight of the graviton field and therefore arbitrary. In (1.2) the value of the real parameter $\bar{c}$ depends essentially on the group representation: For three-dimensional chiral fields, $\bar{c} = 0$ corresponds to a nonet of pseudoscalar mesons, while $\bar{c} = \frac{1}{3}$ corresponds to an octet of such mesons.

[a)]Aangesteld navorser bij het Nationaal Fonds voor Wetenschappelijk Onderzoek (Belgium).

Ashmore and Delbourgo[1,2] were the first to derive analytical expressions in closed form for the gravitational superpropagator $T^{(4)}_{1,1}[c,\Delta]$, and the chiral superpropagator $\bar{T}^{(3)}_{1,1}[\bar{c},\Delta]$. An alternative derivation of the latter result has been given recently by Kapoor.[3] Finally, it has been shown by the present author[4] that the Ashmore–Delbourgo algorithm is also very well suited for the calculation of the multimatrix superpropagators $T^{(\nu)}_{k,1}[c,\Delta]$, $T^{(\nu)}_{k,1}[c,\Delta]$, $\bar{T}^{(\nu)}_{1,k}[\bar{c},\Delta]$, and $\bar{T}^{(\nu)}_{k,1}[\bar{c},\Delta]$ $(\nu\in\mathbb{N}_0,\ k\in\mathbb{N}_0)$.

The exact results obtained show a common property, namely that one can indicate values of c or $\bar{c}$ whereby the superpropagators decrease exponentially in the neighborhood of the singularity of the scalar propagator Δ. Motivated by this observation, Isham, Salam, and Strathdee[5] proposed the conjecture that gravity is possibly ambiguity-free in the sense that all exponential gravitational Green functions might demonstrate such a behavior. In the present paper we shall give a rigorous proof of this conjecture. As a by-product some general properties, also concerning the chiral Green functions, are found.

2. A BASIC THEOREM

As a first step in our proof we state that the superpropagators $T^{(\nu)}_{1,1}[c,\Delta]$ $(\nu\in\mathbb{N}_0)$, of the gravitational kind, depend on the gauge parameter c as given in:

Theorem 1:

$$T^{(\nu)}_{1,1}[c,\Delta] = e^{-\kappa^2 c\Delta}\,T^{(\nu)}_{1,1}[0,\Delta]. \quad (2.1)$$

Proof: For real symmetrical or Hermitian (ν,ν) matrices X and Y, the identity

$$\exp[u(X+Y)] = \exp(uX) + \int_0^u \exp[(u-u_1)X]\,Y\exp[u_1(X+Y)]\,du_1 \quad (u\in\mathbb{R}) \quad (2.2)$$

can be used to deduce by recurrence and a change of variables the series expansion

$$\exp(X+Y) = \sum_{n=1}^{\infty}\int_0^1 dv_1\int_0^1 dv_2\cdots\int_0^1 dv_n\, e^{Xv_1}Ye^{Xv_2}Y\times\cdots Ye^{Xv_n}\,\delta(v_1+v_2+\cdots+v_n-1). \quad (2.3)$$

Equating the matrices X and Y respectively with the diagonal and off-diagonal part of the ϕ matrix, i.e.,

 0022-2488/79/081797-04$01.00

$$\begin{cases} X_{ij} = \kappa\phi_{ii}\,\delta_{ij} \\ \qquad\qquad (i,j = 1,2,...,\nu) \\ Y_{ij} = \kappa\phi_{ij}\,(1-\delta_{ij}) \end{cases} \tag{2.4}$$

and taking then matrix elements on both sides of Eq. (2.3), one obtains the representation

$$\begin{aligned}(e^{\kappa\phi})_{\alpha\beta} = &\sum_{n=1}^{\infty}\sum_{i_1,i_2,...,i_n=1}^{\nu}\delta_{\alpha i_1}\,\delta_{\beta i_n}\,Y_{i_1 i_2}\,Y_{i_2 i_3}\cdots \\ &\times Y_{i_n\ i_1}\int_0^1 dv_1\cdots\int_0^1 dv_n\,\delta(v_1+\cdots+v_n-1) \\ &\times\exp(X_{i_1 i_1}v_1+\cdots+X_{i_n i_n}v_n).\end{aligned} \tag{2.5}$$

Since there are at most ν different diagonal X components in the exponent on the right-hand side of (2.5), some of the variables of integration will not occur explicitly in the exponential; others will stick together as sums. Nevertheless, we can write most generally that

$$\begin{aligned}&\int_0^1 dv_1\cdots\int_0^1 dv_n\,\delta(v_1+\cdots+v_n-1) \\ &\quad\times\exp(X_{i_1 i_1}v_1+\cdots+X_{i_n i_n}v_n) \\ &\quad=\int_0^1 du_1\cdots\int_0^1 du_\nu\,\delta(u_1+\cdots+u_\nu-1)F(u_1,...,u_\nu) \\ &\quad\times\exp\Big(\sum_{i=1}^{\nu}X_{ii}u_i\Big),\end{aligned} \tag{2.6}$$

whereby F is polynomial in the integration variables and whereby one or more of these variables may not occur in the integrand.

Taking traces at both sides of (2.5), one finds with the help of (2.6), that Tr $\exp\kappa\phi$ can be expanded as follows:

$$\begin{aligned}\mathrm{Tr}\exp\kappa\phi = &\sum_{i=1}^{\nu}I_i^{(1)}+\sum_{1\le i<j\le\nu}I_{ij}^{(2)} \\ &+\sum_{1\le i<j<k\le\nu}I_{ijk}^{(3)}+\cdots+I_{12...\nu}^{(\nu)},\end{aligned} \tag{2.7}$$

with

$$I_i^{(1)} = \exp X_{ii}\,,$$

$$\begin{aligned}I_{ij}^{(2)} = &\sum_{\{\alpha_2\}}\Phi^{(2)}[\{\alpha_2\},\{Y_{ij}\}]\int_0^1 du_1\int_0^1 du_2 \\ &\times\delta(u_1+u_2-1)F^{(2)}[\{\alpha_2\},V_2]\exp(X_{ii}u_1+X_{jj}u_2),\end{aligned}$$

$$\begin{aligned}I_{ijk}^{(3)} = &\sum_{\{\alpha_3\}}\Phi^{(3)}[\{\alpha_2\},\{Y_{ij},Y_{jk},Y_{ik}\}]\int_0^1 du_1\int_0^1 du_2 \\ &\times\int_0^1 du_3\,\delta(u_1+u_2+u_3-1)F^{(3)}[\{\alpha_3\},V_3] \\ &\times\exp(X_{ii}u_1+X_{jj}u_2+X_{kk}u_3),\end{aligned} \tag{2.8}$$

$$\begin{aligned}I_{12...\nu}^{(\nu)} = &\sum_{\{\alpha_\nu\}}\Phi^{(\nu)}[\{\alpha_\nu\},\{Y_{ij}\mid i,j\in\mathscr{S}_\nu\,;i\leqslant j\}] \\ &\times\prod_{i=1}^{\nu}\Big(\int_0^1 du_i\Big)\delta\Big(\sum_{i=1}^{\nu}u_i-1\Big)F^{(\nu)}[\{\alpha_\nu\},V_\nu] \\ &\times\exp\Big(\sum_{i=1}^{\nu}X_{ii}u_i\Big),\end{aligned}$$

$$V_k = \{u_1,u_2,...,u_k\}\quad(k\geqslant 2).$$

Herein the sets $\{\alpha_\mu\}$, $(\mu = 2,...,\nu)$ denote minimal sets of independent summation variables needed for counting the powers of the different nondiagonal ϕ elements in the various matrix-product chains of such elements. Defining two chains, containing exactly the same nondiagonal elements (not necessarily in the same order) and therefore also associated wih a same label set $\{\alpha_\mu\}$, equivalent when the values of the corresponding labels are identical, $\Phi^{(\mu)}[\{\alpha_\mu\},\{Y_{\sigma\tau},...,Y_{\lambda\rho}\}]$ stands for a representant of the class formed by the equivalent chains which contain the same elements $Y_{\sigma\tau},...,Y_{\lambda\rho}$, each occuring as many times as indicated by the values of the labels in $\{\alpha_\mu\}$. Finally, the functions $F^{(\mu)}$ are derived from a corresponding function $F(u_1,...,u_\nu)$ in (2.6), by taking into account the combinatorial factors associated with the number of different chains in the class. It is obvious that the task of finding explicit expressions for the latter functions is an enormous one in general. It will become clear, however, that such expressions are not relevant to our proof. Nevertheless, in order to clarify some previous introduced notations, we want to give here exact partial resaults which have been obtained in a different context by Metha and Kumar,[6] namely,

$$\sum_{\{\alpha_2\}}\equiv\sum_{n=1}^{\infty}\,,$$

$$\Phi^{(2)}[n,Y_{ij}]\equiv|Y_{ij}|^{2n}, \tag{2.9}$$

$$F^{(2)}[n,u_1,u_2]\equiv\frac{u_1^{n-1}u_2^{n-1}}{n!(n-1)!}\,,$$

$$\sum_{\{\alpha_3\}}\equiv\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}$$

with $(l+m+p)(m+n+p)(n+l+p)\neq 0$,

$$\begin{aligned}&\Phi^{(3)}[l,m,n,p,Y_{ij},Y_{jk},Y_{ik}] \\ &\quad\equiv|Y_{ij}|^{2l}\,|Y_{jk}|^{2m}\,|Y_{ik}|^{2n}\,[(Y_{ij}Y_{jk}Y_{ki})^p \\ &\qquad+(Y_{ik}Y_{kj}Y_{ji})^p](1-\tfrac{1}{2}\delta_{p,0}),\end{aligned} \tag{2.10}$$

$$\begin{aligned}&F^{(3)}[l,m,n,p,u_1,u_2,u_3] \\ &\quad\equiv\frac{p^2+p(l+m+n)+mn+nl+lm}{(l+p)!(m+p)!(n+p)!l!m!n!} \\ &\qquad\times(u_1u_2)^l(u_2u_3)^m(u_3u_1)^n(u_1u_2u_3)^{p-1}.\end{aligned}$$

Note that the class representatives are written in such a way that the results are applicable both for real symmetrical and for Hermitian matrices Y. Therefore, one has only to choose Y identical to the off-diagonal part of $\kappa\bar{\phi}$ and X identical to the diagonal part of $\kappa\bar{\phi}$ to assure that (2.8) is also a valid formula for the chiral superpropagator case.

The c dependence of the superpropagator $T_{1,1}^{(\nu)}[c,\Delta]$ originates from the vacuum expectation values of functions of the diagonal elements of ϕ only. This follows from the fact that (1.1) may be rewritten as

$$\left.\begin{aligned}&\langle X_{\alpha\alpha},X_{\gamma\gamma}\rangle = \kappa(\delta_{\alpha\gamma}-c)\Delta \\ &\langle X_{\alpha\alpha},Y_{\gamma\delta}\rangle = \langle Y_{\gamma\delta},X_{\alpha\alpha}\rangle = 0 \\ &\langle Y_{\alpha\beta},Y_{\gamma\delta}\rangle = \tfrac{1}{2}\kappa(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma})\Delta\end{aligned}\right\}\quad(\alpha,\beta,\gamma,\delta\in\mathscr{S}_\nu). \tag{2.11}$$

By using now the property that the $I^{(\mu)}$ terms in (2.8) contain the X elements only through factors of the form

$$\prod_{i=1}^{\nu}\Big(\int_0^1 du_i\Big)\delta\Big(\sum_{i=1}^{\nu}u_i-1\Big)F[\{\alpha_\mu\},V_\mu]$$

$$\times\exp\Big(\sum_{i=1}^{\nu} X_{ii}u_i\Big),$$

$(\mu\in\mathcal{J}_\nu)$,

whereby one or more of the integration variables may be absent, and further by using the equalities

$$\Big\langle\exp\Big(\sum_{i=1}^{\nu} X_{ii}u_i\Big),\exp\Big(\sum_{j=1}^{\nu} X_{jj}u'_j\Big)\Big\rangle \times\delta\Big(\sum_{i=1}^{\nu} u_i-1\Big)\delta\Big(\sum_{j=1}^{\nu} u'_j-1\Big)$$
$$=\exp\Big[\kappa^2(1-c)\Delta\sum_{i=1}^{\nu} u_iu'_i-\kappa^2c\Delta\sum_{\substack{i,j=1\\(i\neq j)}}^{\nu} u_iu'_j\Big]\times\delta\Big(\sum_{i=1}^{\nu} u_i-1\Big)\delta\Big(\sum_{j=1}^{\nu} u'_j-1\Big)$$
$$=\exp\Big(-\kappa^2c\Delta+\kappa^2\Delta\sum_{i=1}^{\nu} u_iu'_i\Big)\times\delta\Big(\sum_{i=1}^{\nu} u_i-1\Big)\delta\Big(\sum_{j=1}^{\nu} u'_j-1\Big) \tag{2.12}$$

which are easily deduced with the help of (2.11) and which are independent of the number of integration variables u and u' present, the substitution of (2.8) into (2.7) immediately leads to the result (2.1) of Theorem 1.

It has to be noted that, in the present particular case, the contributions coming from vacuum expectation values of two $I^{(\mu)}$ terms which do not have exactly the same number of indices with equal values disappear as a consequence of (2.11). Finally, we note that we have not found any indication of the validity of Theorem 1 by analyzing the Ashmore–Delbourgo algorithm, and this is a fortiori the case for the generalizations of the next section, which concern Green functions wherefore the algorithm cannot even be invoked for calculating exact expressions.

3. EXTENSIONS

Having remarked already that (2.7) and (2.8) remain valid formulas by setting

$$\left.\begin{aligned}X_{ij}&=\kappa\bar{\phi}_{ij}\delta_{ij}\\Y_{ij}&=\kappa\bar{\phi}_{ij}(1-\delta_{ij})\end{aligned}\right\}\quad(i,j=1,2,\ldots,\nu), \tag{3.1}$$

It suffices to proceed as in (2.12), with c replaced by $\bar{c}$, to prove the following:

Theorem 2:

$$\bar{T}^{(\nu)}_{1,1}[\bar{c},\Delta\,]=e^{\kappa^2\bar{c}\Delta}\,\bar{T}^{(\nu)}_{1,1}[0,\Delta\,]. \tag{3.2}$$

Next, we state that the previous theorems can be extended to more general Green functions as follows:

Theorem 3:

$$T^{(\nu)}_{k_1,\ldots,k_n}[c,\{\Delta_{ij}\mid i,j\in\mathcal{J}_n;i<j\}]=\exp\Big(-\kappa^2c\sum_{i<j=1}^{n}k_ik_j\Delta_{ij}\Big)T^{(\nu)}_{k_1,\ldots,k_n}[0,\{\Delta_{ij}\mid i,j\in\mathcal{J}_n;i<j\}], \tag{3.3}$$

and

Theorem 4:

$$\bar{T}^{(\nu)}_{k_1,\ldots,k_n}[\bar{c},\{\Delta_{ij}\mid i,j\in\mathcal{J}_n;i<j\}]=\exp\Big(-\kappa^2\bar{c}\sum_{i<j=1}^{n}k_ik_j\Delta_{ij}\Big)\bar{T}^{(\nu)}_{k_1,\ldots,k_n}[0,\{\Delta_{ij}\mid i,j\in\mathcal{J}_n;i<j\}]. \tag{3.4}$$

We note that the latter theorems generalize the former ones in two ways: Not only can there be more than one traced exponential of the matrix field in a certain space-time point, but also the number of different space-time points may be greater than one.

The proof of (3.2) and equivalently of (3.3) is again based on the representation (2.7), (2.8), and follows exactly the same pattern as before. In fact, the essential step here is the derivation of the equalities:

$$\Big\langle\Big[\exp\Big(\sum_{i=1}^{\nu}X_{ii}u_i^{(1)}\Big)\Big]^{k_1},\Big[\exp\Big(\sum_{j=1}^{\nu}X_{jj}u_j^{(2)}\Big)\Big]^{k_2},\ldots,\Big[\exp\Big(\sum_{l=1}^{\nu}X_{ll}u_l^{(n)}\Big)\Big]^{k_n}\Big\rangle\prod_{s=1}^{n}\delta\Big(\sum_{i=1}^{\nu}u_i^{(s)}-1\Big)$$
$$=\Big\{\prod_{p<q=1}^{\nu}\exp\Big[\kappa^2(1-c)k_pk_q\Delta_{pq}\sum_{i=1}^{\nu}u_i^{(p)}u_i^{(q)}-\kappa^2ck_pk_q\Delta_{pq}\sum_{\substack{i,j=1\\(i\neq j)}}^{\nu}u_i^{(p)}u_j^{(q)}\Big]\Big\}\prod_{s=1}^{n}\delta\Big(\sum_{i=1}^{\nu}u_i^{(s)}-1\Big)$$
$$=\Big\{\prod_{p<q=1}^{\nu}\exp\Big(-\kappa^2ck_pk_q\Delta_{pq}+\kappa^2k_pk_q\Delta_{pq}\sum_{i=1}^{\nu}u_i^{(p)}u_j^{(q)}\Big)\Big\}\prod_{s=1}^{n}\delta\Big(\sum_{i=1}^{\nu}u_i^{(s)}-1\Big), \tag{3.5}$$

which, after replacing c by $\bar{c}$, also hold for the chiral case, i.e., for X and Y defined as in (3.1).

4. PROOF OF THE ISHAM CONJECTURE

Since we know the explicit c dependence ($\bar{c}$ dependence) of the various exponential gravity (chiral) Green functions, we can now set without loss of generality c ($\bar{c}$) equal to zero everywhere. Then it follows from (1.1) [(1.2)] that the functions $T^{(\nu)}_{k_1,\ldots,k_n}[0,\{\Delta_{ij}\mid i,j\in\mathcal{J}_n;i<j\}]$ ($\bar{T}^{(\nu)}_{k_1,\ldots,k_n}[0,\{\Delta_{ij}\mid i,j\in\mathcal{J}_n;i<j\}]$) are the series sums of nonnegative terms. In any case, a useful upper bound is given by the corresponding traced Green functions associated with a (ν,ν)-matrix field

$\psi_{\alpha\beta}(x)$ propagating as

$$\langle \psi_{\alpha\beta}(x), \psi_{\gamma\delta}(y) \rangle = \Delta(x-y) \quad (\alpha,\beta,\gamma,\delta \in \mathcal{J}_\nu).$$

The calculation of these functions can most easily be performed by replacing consistently $\mathrm{Tr}\psi(x_i)^n$ by φ_i^n, wherby $\varphi_i \equiv \varphi(x_i)$ is a scalar field which on its turn propagates as

$$\langle \varphi_i, \varphi_j \rangle = \nu^2 \Delta_{ij}. \tag{4.1}$$

With the help of the well-known result

$$\langle (\exp\varphi_1)^{k_1}, (\exp\varphi_2)^{k_2}, \ldots, (\exp\varphi_n)^{k_n} \rangle = \exp\left(\sum_{i<j=1}^{n} k_i k_j \langle \varphi_i, \varphi_j \rangle \right), \tag{4.2}$$

we easily obtain as a corollary of Theorem 3 that

$$T^{(\nu)}_{k_1,\ldots,k_n}[c, \{\Delta_{ij} \mid i,j \in \mathcal{J}_n; i<j\}] \leqslant \exp\left[\kappa^2(\nu^2 - c) \sum_{i<j=1}^{\nu} k_i k_j \Delta_{ij} \right], \tag{4.3}$$

the equality being attained only if $\nu = 1$.

By choosing c only for all times greater than ν^2 it is thus guaranteed that all gravity exponential Green functions are asymptotically (this means when one of the Δ_{ij} functions tends to $+\infty$) exponentially decreasing, which proves the Isham conjecture.[5] It is obvious from the foregoing that we cannot claim to have found the least lower bound on the c values which assure exponential decrease for large values of one of the Δ_{ij} functions. Indeed, by analyzing the exact asymptotic behavior of a variety of traced multimatrix superpropagators given in an earlier contribution,[7] we have good reasons to believe that $1+\epsilon$, with $\epsilon > 0$ and arbitrary small, might as well be a "good" lower bound. Finally, we can write a formula analogous to (4.3) for $\bar{T}^{(\nu)}_{k_1,\ldots,k_n}[\bar{c}, \{\Delta_{ij} \mid i,j \in \mathcal{J}_n; i<j\}]$, but since $\bar{c}$ is not arbitrary, this does not lead us, in general, to suppression of ultraviolet infinities in exponentially parametrized chiral field theories.

ACKNOWLEDGMENTS

We would like to thank Professor C.C. Grosjean for his continuous interest during the course of this work.

[1] J. Ashmore and R. Delbourgo, J. Math. Phys. **14**, 176 (1973).
[2] J. Ashmore and R. Delbourgo, J. Math. Phys. **14**, 569 (1973).
[3] A.K. Kapoor, J. Math. Phys. **17**, 61 (1976).
[4] H.E. De Meyer, J. Phys. A **9**, 1333 (1976).
[5] C.J. Isham, A. Salam, and J. Strathdee, Phys. Lett. B **46**, 407 (1973).
[6] M.L. Metha and K. Kumar, J. Phys. A **9**, 197 (1976).
[7] H.E. De Meyer, J. Phys. A **11**, 885 (1978).

Structure and representation of correlation functions and the density matrix for a statistical wave field in optics

E. C. G. Sudarshan

Center for Particle Theory, University of Texas, Austin, Texas 78712

N. Mukunda

Centre for Theoretical Studies, Indian Institute of Science, Bangalore, India 560 012

(Received 20 October 1978; revised manuscript received 13 November 1978)

A systematic structure analysis of the correlation functions of statistical quantum optics is carried out. From a suitably defined auxiliary two-point function we are able to identify the excited modes in the wave field. The relative simplicity of the higher order correlation functions emerge as a byproduct and the conditions under which these are mode pure are derived. These results depend in a crucial manner on the notion of coherence indices and of unimodular coherence indices. A new class of approximate expressions for the density operator of a statistical wave field is worked out based on discrete characteristic sets. These are even more economical than the diagonal coherent state representations. An appreciation of the subtleties of quantum theory obtains. Certain implications for the physics of light beams are cited.

INTRODUCTION

Light is essentially quantum mechanical in nature.[1] A light beam is therefore to be represented by an ensemble of quantum states. For a variety of purposes it is convenient to specify a statistical state by giving the set of correlation functions of all possible orders corresponding to that state. It is therefore of value to study representations of the statistical state that make the calculation of correlation functions as direct as possible. One must of course satisfy oneself that a set of (acceptable) correlation functions defines the state uniquely.

By a systematic analysis of the structure and properties of the correlation functions we can identify the "modes," i.e., the natural one-photon wavefunctions, in terms of which the given correlation function, and so the associated states, assume their simplest form. Such an analysis basically brings out the consequences of the positivity properties of the density matrix and the Bose nature of light.[2] Unlike the case, say, of hydrodynamic turbulence,[3] the correlation functions for light obey uncoupled equations of propagation,[4] and the interaction with matter can be treated perturbatively. This fact justifies the structure analysis of statistical states of the free electromagnetic field and also leads to a remarkable simplification in that we basically need study only systems with one degree of freedom.

Correlation functions for the quantized free electromagnetic field are defined as expectation values of normal ordered operator functions of the field.[5] The diagonal coherent state representation[6] of the density matrix therefore leads to expressions for these functions closely mimicking classical correlation functions defined as averages over classical statistical ensembles.[4] The diagonal representation asserts that any density operator for a quantum system with one degree of freedom may be displayed in the form

$$\rho = \int \phi(z)|z\rangle\langle z|d\mu(z),$$

$$d\mu(z) = (1/\pi)d^2z = (1/\pi)d(\mathrm{Re}z)d(\mathrm{Im}z).$$

Here ϕ is a c-number weight function, the coherent states $|z\rangle$ are eigenstates of the annihilation operator for complex eigenvalues z and the integration is over the entire complex plane.[6] These states taken for all z form at the vector space level an overcomplete family[7] permitting a general vector to be expanded linearly in terms of them in more than one way. The diagonal representation for statistical states exploits this overcompleteness to avoid all nondiagonal outer products of coherent states in the above representation and so achieve a form similar to a classical ensemble. The weight function ϕ is generally a distribution belonging to the space Z_2. Nevertheless with this weight function all "normal ordered" correlation functions can be calculated as if the system were classical with the same weights.[6] It must be stressed that we have here a description of the fully quantum mechanical system and not just of its classical limit or semiclassical approximation. The quantum nature of the system lies in the set of weights ϕ to be admitted, and in their properties. Such a description is valuable in any assessment of the information carried by a light beam.

For such a diagonal representation to exist it is necessary that the set of states used be overcomplete. A set of states constituting a complete, but not overcomplete, basis at the vector space level would represent almost all density operators in a nondiagonal form. Since a system with one degree of freedom already calls for a countably infinite set of basis vectors, we do expect every complete as well as overcomplete set to contain infinitely many elements. The coherent states form a two-parameter continuous infinity of states labelled by the real and imaginary parts of z. The question naturally arises as to whether there are "smaller" overcom-

 0022-2488/79/081801-10$01.00

plete families of states which are yet rich enough to permit a diagonal representation for any density operator in terms of them. More precisely, are there smaller families using which we can approximate any given density operator through expressions of the diagonal form to any desired accuracy?

We find that there are indeed such families which need not even be continuously infinite. There is a great variety of countable overcomplete sets of coherent states using any one of which arbitrarily good diagonal approximations to a given density operator can be obtained.[8] It is curious, however, that if we spread these states "as uniformly as possible" over the complex z-plane, then these states must be more dense than one per unit phase cell in most parts of the phase plane. Indeed, Planck's constant fails to provide a natural size for cells in phase space with respect to which the density of such overcomplete sets may be meaningfully stated. This indicates yet another subtle aspect of quantum theory.

The picture of the general density operator for a light beam has then the following structure: The (electric) field is expanded in terms of a complete orthonormal set of natural modes. Under rather general conditions this is a discrete set. All these natural modes will generally contribute to the two-point correlation function. The density operator can now be associated with a multivariate weight function in the excitations of these modes. If we use all the coherent states associated with the annihilation operator corresponding to each mode, we deal with the diagonal representation in its conventional form, and the weight function is in general a distribution in a discrete set of complex variables one per mode. We may alternatively choose a countable overcomplete set of coherent states to go with each mode, rich enough to allow diagonal-type approximations to any statistical state as far as this mode is concerned; thus any state with respect to this mode can be approximated through ensembles over the chosen discrete set of complex eigenvalues for the mode annihilation operator. And a general density operator for the total field system can be approximated arbitrarily closely by ensembles over the collection of discrete sets of eigenvlaues for the annihilation operators of all modes.

The plan of the paper is as follows. Section 1 gives a brief development of the quantum mechanics of a system with one degree of freedom mostly with a view to establishing notation and deriving certain results in a form to be used later. Sections 2 and 3 deal with the analysis of correlation functions and the unravelling of natural modes; both sections are primarily concerned with the descriptions of those parts of a statistical state that correspond to "large photon numbers." After preliminaries, Sec. 2 analyzes this aspect for a state for which it is assumed that a correlation function of some definite order obeys a *condition of coherence*. Section 3 on the other hand analyzes this aspect for a general state by a slightly different means. Section 4 introduces the Weyl operators and the associated expansions of density operators. The problem of the diagonal representation for operators is then identified with the one representing an arbitrary vector state in terms of linear combinations of subsets of coherent states. The basic tool here is the notion of a second Hilbert space made up of operators on the original one. In Sec. 5 we make use of these results and those given in Sec. 1 to derive the general discrete-discrete approximation to the density operator of a light beam. Concluding remarks and some open problems make up Sec. 6.

1. QUANTUM THEORY OF ONE DEGREE OF FREEDOM

A classical one-dimensional harmonic oscillator with unit mass and frequency ω has canonical variables q, p obeying equations of motion

$$\dot{q} = p, \quad \dot{p} = -\omega^2 q. \tag{1.1}$$

Use of the complex canonical variables

$$a = (\omega q + ip)/\sqrt{2\omega}, \quad a^* = (\omega q - ip)/\sqrt{2\omega}, \tag{1.2}$$

allows the solution of Eq. (1.1) to be completely expressed as

$$a(t) = a(0)\exp(-i\omega t), \quad a^*(t) = a^*(0)\exp(i\omega t). \tag{1.3}$$

The instantaneous state of the oscillator is given by the value of the complex dynamical variable a, and as time advances the representative point in the a plane describes a circle. This plane is a rescaled version of the phase space.

For a quantum system with one degree of freedom we have two unbounded Hermitian operators q, p obeying the commutation relation

$$[q,p] = i. \tag{1.4}$$

(Planck's constant has been set equal to unity, and no special symbols such as carets are used to distinguish operators, as there will be no cause for confusion on this account.) This relation can be transcribed in terms of bounded operators by introducing the Weyl families of unitary operators

$$U(\sigma) = \exp(i\sigma q), \quad V(\tau) = \exp(i\tau p), \qquad -\infty < \sigma,\tau < \infty, \tag{1.5}$$

Then Eq. (1.4) is equivalent to

$$\begin{aligned} &U(\sigma)U(\sigma') = U(\sigma+\sigma'),\\ &V(\tau)V(\tau') = V(\tau+\tau'),\\ &U(\sigma)V(\tau) = V(\tau)U(\sigma)\exp(-i\sigma\tau). \end{aligned} \tag{1.6}$$

Setting the frequency ω equal to unity for simplicity, the annihilation operator a and its Hermitian adjoint $a^\dagger$ are defined, following the classical definition (1.2), as

$$a = (q + ip)/\sqrt{2}, \quad a^\dagger = (q - ip)/\sqrt{2}, \tag{1.7}$$

and then the commutation relation (1.4) appears as

$$\{a,a^\dagger\} = 1. \tag{1.8}$$

Coherent states[9] are eigenvectors of the annihilation operator,

$$a|z\rangle = z|z\rangle, \tag{1.9}$$

with the eigenvalue z being any complex number. These states are normalizable, and when normalized to unity their Schrödinger wavefunctions may be taken to be

$$\langle q'|z\rangle = \pi^{-1/4}\exp\left[-\tfrac{1}{2}(q' - z\sqrt{2})^2 - \tfrac{1}{2}z(z^* - z)\right]. \tag{1.10}$$

No two of these states are mutually orthogonal as one has

$$\langle z|z'\rangle = \exp(-\tfrac{1}{2}|z|^2 - \tfrac{1}{2}|z'|^2 + z^*z'). \tag{1.11}$$

The coherent states taken together for all complex z are complete as they furnish a resolution of the identity in the form[10]

$$\psi = \int |z\rangle\langle z|d\mu(z) \equiv (1/\pi)\int |z\rangle\langle z|d^2z. \tag{1.12}$$

Actually, however, they are overcomplete, as one can easily exhibit linear dependences among them in the form of integral relationships. On the other hand any finite number of distinct coherent states are linearly independent. We come back to the use of Eq. (1.12) in a moment.

We can rewrite the operators (1.5) of the Weyl family in the form

$$U(\sigma) = \exp[i\sigma(a + a^\dagger)/\sqrt{2}],$$
$$V(\tau) = \exp[\tau(a - a^\dagger)/\sqrt{2}]. \tag{1.13}$$

This motivates the introduction of the more general Weyl family of unitary operators

$$W(\alpha) = \exp(\alpha a^\dagger - \alpha^* a)$$
$$= U\left(\frac{\alpha - \alpha^*}{i\sqrt{2}}\right) V\left(-\frac{\alpha + \alpha^*}{\sqrt{2}}\right)$$
$$\times \exp\left(\frac{\alpha^{*2} - \alpha^2}{4}\right), \tag{1.14}$$

where α is any complex number. This family will be put to use in Sec. 4. At this point we note the diagonal coherent state matrix elements of these operators,

$$\langle z|W(\alpha)|z\rangle = \exp(-\tfrac{1}{2}|\alpha|^2 + \alpha z^* - \alpha^* z). \tag{1.15}$$

Because of the overcompleteness of coherent states, one expects to be able to "expand" any state $|\psi\rangle$ in terms of them in more than one way. A particular expansion is supplied by the resolution of the identity, Eq. (1.12); one has for any $|\psi\rangle$,

$$|\psi\rangle = \int \langle z|\psi\rangle |z\rangle\, d\mu(z). \tag{1.16}$$

The particular "wavefunction" occurring in this expansion has certain characteristic features. If for convenience we write

$$\langle z^*|\psi\rangle = \exp(-\tfrac{1}{2}|z|^2) f(z), \tag{1.17}$$

then $f(z)$ is an entire analytic function whose behavior for large $|z|$ is controlled by

$$|f(z)| \leqslant \| |\psi\rangle \| \exp(\tfrac{1}{2}|z|^2). \tag{1.18}$$

Using Eq. (1.10) we can relate $f(z)$ to the Schrödinger wavefunction of $|\psi\rangle$ through

$$f(z) = \pi^{-1/4} \exp(-\tfrac{1}{2}z^2) \int_{-\infty}^{\infty} \psi(q')$$
$$\times \exp(-\tfrac{1}{2}q'^2 + q'z\sqrt{2})\, dq'. \tag{1.19}$$

An alternative expansion possibility arises by considering a suitable subset of coherent states. One choice of subset is given by $|i\sqrt{2}r\rangle$ for all real r. If we tentatively write an expansion

$$|\psi\rangle = \pi^{-1/2} \int_{-\infty}^{\infty} v(r)|i\sqrt{2}r\rangle\, dr, \tag{1.20}$$

a means must be found to evaluate the weight function $v(r)$. One way is to take the scalar product of both sides of Eq. (1.20) with another vector of the subset, $|i\sqrt{2}r'\rangle$. Then use of Eq. (1.11), (1.17), (1.19) gives:

$$\int_{-\infty}^{\infty} v(r) \exp[-(r - r')^2]\, dr$$
$$= \pi^{1/4} \int_{-\infty}^{\infty} \psi(q') \exp(-\tfrac{1}{2}q'^2 + 2ir'q')\, dq', \tag{1.21}$$

so that

$$v(r) = \pi^{-1/4} \int_{-\infty}^{\infty} \psi(q') \exp(\tfrac{1}{2}q'^2) \exp(2irq')\, dq'. \tag{1.22}$$

Since the integrand may grow fast at infinity the weight function $v(r)$ may not be an ordinary function but a distribution (in the family Z_2). Alternately we may take the scalar product of the two sides of Eq. (1.20) with a coherent state $|\sqrt{2}s\rangle$, s real, to obtain

$$\langle\sqrt{2}s|\psi\rangle = \pi^{-1/2} \int_{-\infty}^{\infty} v(r) \exp(-s^2 - r^2 + 2irs)\, dr,$$

so that

$$v(r) = \frac{\exp(r^2)}{\sqrt{\pi}} \int_{-\infty}^{\infty} \langle\sqrt{2}s|\psi\rangle \exp(s^2 - 2irs)\, ds. \tag{1.23}$$

The two possible expansions (1.16, 1.20), with characteristically different properties for their integrands, use, respectively, a two-parameter and a one-parameter continuous infinity of coherent states. Instead of such expansions, if we were satisfied with merely being able to approximate arbitrarily closely to any $|\psi\rangle$ through combinations of coherent states, more economical possibilities in terms of the so-called characteristic sets exist.[11] A set S of points in the complex plane is a characteristic set if we can assert that

$$\langle z|\psi\rangle = 0, \quad z \in S \Rightarrow |\psi\rangle = 0. \tag{1.24}$$

A set S with a finite limit point; the set of all real numbers; the set of all imaginary numbers; any sequence $\{z_n\}$ of distinct nonzero complex numbers for which

$$\sum_{n=1}^{\infty} |z_n|^{-2-\epsilon} = \infty \tag{1.25}$$

for some positive ϵ—all these are examples of characteristic sets. Let us restrict ourselves to discrete sets. In terms of the corresponding coherent states $|z_n|$ we could approximate a given $|\psi\rangle$ to arbitrary accuracy: For each $\eta > 0$ we can find an integer $N(\eta)$ and coefficients $b_n(\eta)$ such that

$$\left\| |\psi\rangle - \sum_{n=1}^{N(\eta)} b_n(\eta)|z_n\rangle \right\| < \eta. \tag{1.26}$$

However, for a general vector $|\psi\rangle$ there is no guarantee that there exists a definite set of coefficients b_n such that the sequence of vectors

$$|\psi_N\rangle = \sum_{n=1}^{N} b_n|z_n\rangle$$

forms a Cauchy sequence converging to $|\psi\rangle$.

We conclude this resumé of the properties of coherent states by quoting one more interesting example of a characteristic set and mentioning a property of such sets in general. The example, due to von Neumann and Perelomov,[12] is the set made up of points in the complex plane of the form

$$z = \sqrt{s}(l + im), \quad l,m = 0, \pm 1, \pm 2, \cdots \quad ; s < \pi. \tag{1.27}$$

Thus by essentially picking one coherent state in a phase volume smaller than a unit cell in phase space we get a set with which any $|\psi\rangle$ can be approximated arbitrarily closely. The property of characteristic sets we have in mind is this: if from any such set S any finite number of points are removed, the remaining points still make up a characteristic set.

2. CORRELATION FUNCTIONS AND NATURAL MODES

The free electromagnetic field may be characterized either by the transverse vector potential $\mathbf{A}(\mathbf{r},\mathbf{t})$ or by the electric field $\mathbf{E}(\mathbf{r},\mathbf{t})$, both of which are transverse and gauge invariant. The interaction Hamiltonian of an electron with the field[13] is expressed in terms of $\mathbf{A}$ directly, so that theoretical expressions relating to experiments based on photoelectric detection naturally involve $\mathbf{A}$ also.[14] We shall thus choose $\mathbf{A}$ as the basic variable for defining correlation functions, though one can always pass to the variable $\mathbf{E}$ by time differentiation. The positive frequency part of $\mathbf{A}(\mathbf{r},\mathbf{t})$, sometimes called the analytic signal,[15,4] consists entirely of annihilation operators and has the time dependence

$$\mathbf{A}^{(+)}(\mathbf{r},t) = e^{-i\hat{\omega}t}\mathbf{A}^{(+)}(\mathbf{r},0), \tag{2.1}$$

where $\hat{\omega}$ is the (positive) frequency operator defined by the wave equation

$$(\nabla^2 + \hat{\omega}^2)\mathbf{A}^{(+)}(\mathbf{r},0) = 0, \quad \hat{\omega} = (-\nabla^2)^{1/2}. \tag{2.2}$$

The transversality of $\mathbf{A}^{(+)}$ is expressed by

$$\nabla\cdot\mathbf{A}^{(+)}(\mathbf{r},t) = 0. \tag{2.3}$$

Therefore, there are only two independent components to $\mathbf{A}^{(+)}$. In momentum space these are the components orthogonal to the momentum direction, and may be chosen to be the two circular polarizations denoted by a two-valued polarization index ϵ. Therefore, we may write for the vector potential,

$$\mathbf{A}^{(+)}(\mathbf{r},t) \to V(\mathbf{r},\epsilon,t) \equiv V(x,t). \tag{2.4}$$

When it is not essential to indicate the polarization index ϵ explicitly, it will be combined with the position vector $\mathbf{r}$ into a single symbol x; and formal integration over x will mean a sum over ϵ plus ordinary integration over space. The time t is *not* combined into x in this way. In fact in all the following analysis we shall be concerned only with conditions at one instant of time, and time variables t will be dropped entirely.

Let some statistical state of the field be given, and let the corresponding density operator be ρ. The general (m,n) order correlation function is defined as the expectation value of the normal-ordered product[5] of m negative frequency (creation) field operators and n positive frequency (annihilation) field operators:

$$\begin{aligned}\Gamma^{(m,n)}(x_1,x_2,\ldots,x_m;y_1,y_2,\ldots,y_n) &= \langle V(x_1)^\dagger\cdots V(x_m)^\dagger V(y_1)\cdots V(y_n)\rangle \\ &= \mathrm{Tr}[V(y_1)\cdots V(y_n)\rho V(x_1)^\dagger\cdots V(x_m)^\dagger].\end{aligned} \tag{2.5}$$

Here, m and n are nonnegative integers, and for given values of these, the correlation function depends symmetrically on the m x's and also on the n y's. For conciseness, we may write $\Gamma^{(m,n)}(x;y)$ for the above correlation function; the superscripts imply that x actually stands for m arguments, each consisting of a position vector and a polarization label, etc. As seen by inspection of Eq. (2.5), the following relation holds,

$$\Gamma^{(n,m)}(y,x) = [\Gamma^{(m,n)}(x;y)]^*. \tag{2.6}$$

Further, for $m = n$, the "matrix" $\Gamma^{(m,m)}(x;y)$ with continuous matrix indices x and y is nonnegative,

$$\int dx_1\cdots dx_m\, dy_1\cdots dy_m f(x_1\cdots x_m)^*\Gamma^{(m,m)}(x;y)f(y_1\cdots y_m) \geqslant 0. \tag{2.7}$$

More generally, we may view the entire collection $\Gamma^{(m,n)}(x;y)$ as constituting a giant matrix Γ in which $\Gamma^{(m,n)}$ stands in the (m,n) position when Γ is partitioned.[16] If similarly f is a giant vector at whose mth position stands a symmetric function $f_m(x_1,\ldots,x_n)$ of m x's, then we may identify Γ as a nonnegative matrix according to

$$f^\dagger\Gamma f \equiv \sum_{m,n=0}^{\infty}\int dx_1\cdots dx_m dy_1\cdots dy_n f_m(x_1\cdots x_m) \times^*\Gamma^{(m,n)}(x;y)f_n(y_1\cdots y_n) \geqslant 0. \tag{2.8}$$

We now wish to study in more detail some general properties of these correlation functions, especially for "large" orders.

For this purpose we introduce the family of "optical discriminants" $\Delta^{(m,n)}(x;y)$ according to

$$\Delta^{(m,n)}(x;y) \equiv \Gamma^{(m,m)}\Gamma^{(n,n)}(y;y) - \Gamma^{(m,n)}(x;y)\,\Gamma^{(n,m)}(y;x). \tag{2.9}$$

These are real functions of the indicated variables, and the essential point in the definition is that these are nonnegative quantities. In fact, in terms of the operator combination

$$G^{(m,n)}(x;y) = \Gamma^{(n,n)}(y;y)V(x_1)\cdots V(x_m) - \Gamma^{(n,m)}(y;x)V(y_1)\cdots V(y_n), \tag{2.10}$$

we have

$$\begin{aligned}&\Delta^{(m,n)}(x;y)\Gamma^{(n,n)}(y;y) \\ &\quad = \mathrm{Tr}(G^{(m,n)}(x;y)\rho G^{(m,n)}(x;y)^\dagger \geqslant 0,\end{aligned} \tag{2.11}$$

so that (provided $\Gamma^{(n,n)}$ does not vanish identically) it follows that

$$\Delta^{(m,n)}(x;y) \geqslant 0. \tag{2.12}$$

Another expression of this result can be given in terms of the "coherence indices." The coherence index of order (m,n) is defined to be

$$S^{(m,n)}(x;y) = \Gamma^{(m,n)}(x;y)/[\Gamma^{(m,m)}(x;x)\Gamma^{(n,n)}(y;y)]^{1/2}. \quad (2.13)$$

Then Eq. (2.12) says the same thing as

$$0 \leqslant |S^{(m,n)}(x;y)| \leqslant 1. \quad (2.14)$$

Let us now have a statistical state ρ for which it is given that for a certain pair of integers (m,n), the inequality in Eq. (2.12) becomes an equality for all choices of x and y. (We assume that $\Gamma^{(n,n)}$ is not identically vanishing, and without loss of generality set $m \geqslant n$.) Equivalently, ρ is such that the coherence index of order (m,n) is unimodular for all arguments:

$$|S^{(m,n)}(x;y)| = 1. \quad (2.15)$$

Two interesting questions arise: (i) what can be said about the other correlation functions $\Gamma^{(m',n')}$; (ii) what is the form of the most general ρ that leads to Eq. (2.15)? We examine these in turn, basing the analysis just on positivity of ρ and the Bose nature of V. The characteristic differences between the cases $m = n$ and $m > n$ will be pointed out at the appropriate places.

As pointed out elsewhere[16] that the coherence indices have a maximum modulus of unity may be seen as a consequence of Schwarz's inequality as applied to the m-fold and n-fold products of the field operator. The coherence index may be viewed as a generalized visibility index. In the analysis of Ref. 16 it was concluded that unimodularity of $A\ (n,n)$ implies it for all $S(m',n')$ with $\max(m',n') \geqslant n$. We shall see below that an even stronger conclusion can be derived from the unimodularity of $S(m,n)$ with $m \neq n$.

The vanishing of $\Delta^{(m,n)}(x;y)$, combined with Eq. (2.11) and the fact that ρ is nonnegative, leads to the following operator condition on ρ:

$$V(x_1)\cdots V(x_m)\rho = [\Gamma^{(n,m)}(y;x)/\Gamma^{(n,n)}(y;y)]\, V(y_1)\cdots V(y_n)\,\rho. \quad (2.16)$$

This is to be satisfied for all independent choices of the x's and y's. This is easily exploited to give the relations

$$\Gamma^{(m,n)}(x;y) = \Gamma^{(m,n)}(x;y')\Gamma^{(n,n)}(y';y)/\Gamma^{(n,n)}(y';y')$$

$$= \Gamma^{(m,m)}(x;x')\Gamma^{(m,n)}(x';y)/\Gamma^{(m,m)}(x';x'), \quad (2.17)$$

where again all arguments may be chosen independently. Stated in terms of the coherence indices, these relations are

$$S^{(m,n)}(x;y) = S^{(m,n)}(x;y')\, S^{(n,n)}(y';y)$$

$$= S^{(m,m)}(x;x')\, S^{(m,n)}(x';y). \quad (2.18)$$

In boths Eqs. (2.17) and (2.18) we have two independent relations only if $m > n$. These functional equations imply that each of the three coherence indices has a separable dependence on its two sets of arguments:

$$S^{(m,n)}(x;y) = S^{(m)}(x)/S^{(n)}(y),$$
$$S^{(m,m)}(x;x') = S^{(m)}(x)/S^{(m)}(x'),$$
$$S^{(n,n)}(y;y') = S^{(n)}(y)/S^{(n)}(y'), \quad (2.19)$$

We have here a symmetric function $S^{(m)}$ of m arguments and another $S^{(n)}$ of n arguments, the two coinciding for $m = n$. Then Eq. (2.15) shows that $S^{(m)}$ and $S^{(n)}$ may each be taken to be unimodular, which in turn leads to both $S^{(m,m)}$ and $S^{(n,n)}$ being unimodular along with $S^{(m,n)}$.

The result (2.19) can now be fed into the operator condition (2.16) on ρ, and this then takes the form

$$V(x_1)\cdots V(x_m)\rho = [\alpha^{(m)}(x)/\alpha^{(n)}(y)]\, V(y_1)\cdots V(y_n)\,\rho,$$
$$\alpha^{(m)}(x) \equiv [\Gamma^{(m,m)}(x,x)]^{1/2}/S^{(m)}(x),$$
$$\alpha^{(n)}(y) \equiv [\Gamma^{(n,n)}(y,y)]^{1/2}/S^{(n)}(y). \quad (2.20)$$

Once again we note that we have two distinct symmetric functions $\alpha^{(m)}$ and $\alpha^{(n)}$ when $m > n$ and just one when $m = n$, and the operator condition holds for all choices of x's and y's. We now exploit the fact that any two V's commute:

$$V(\xi)V(x_1)\cdots V(x_m)\rho = V(x_1)V(\xi)\cdots V(x_m)\rho,$$

i.e.,

$$V(\xi)[V(y_1)\cdots V(y_n)/\alpha^{(n)}(y)]\rho$$
$$= [\alpha^{(m)}(\xi,x_2,\ldots,x_m)/\alpha^{(m)}(x_1,x_2,\ldots,x_m)]$$
$$\times V(x_1)[V(y_1)\cdots V(y_n)/\alpha^{(n)}(y)]\,\rho. \quad (2.21)$$

Since the right-hand side must be independent of $x_n,\ldots,x_m$, it follows in the first instance that the dependence of $\alpha^{(m)}$ on its first argument must separate from its dependence on the remaining ones; but since $\alpha^{(m)}$ is a symmetric function, it must factorize all the way. Thus we get the result[17]

$$\alpha^{(m)}(x) = C_m\, u(x_1)\cdots u(x_m), \quad (2.22)$$

where we assume $u(x)$ is a normalized "mode" function, and C_m is a constant. By a parallel argument, for $m > n$, we have

$$\alpha^{(n)}(y) = C_n\, v(y_1)\cdots v(y_n), \quad (2.23)$$

but it is easy to show that v must coincide with u. Assuming this done, the operator condition (2.16), or equally well (2.20), on ρ has the form

$$V(x_1)\cdots V(x_m)\,\rho = [C_m\, u(x_1)\cdots u(x_m)/C_n u(y_1)\cdots u(y_n)]$$
$$\times V(y_1)\cdots V(y_n)\,\rho. \quad (2.24)$$

To fully exploit the fact that this must hold for all x and y, let us introduce the annihilation operator a for the mode u, and its adjoint as

$$a = \int (dx)\, u^*(x)V(x), \qquad a^\dagger = \int dx\, u(x)V(x)^\dagger. \quad (2.25)$$

Then Eq. (2.24) is equivalent to three conditions on ρ:

$$V(x_1)\cdots V(x_m)\,\rho = u(x_1)\cdots u(x_m)\, a^m\rho, \quad (2.26a)$$
$$V(y_1)\cdots V(y_n)\,\rho = u(y_1)\cdots u(y_n)\, a^n\rho, \quad (2.26b)$$
$$a^m\rho = (C_m/C_n)a^n\rho. \quad (2.26c)$$

For $m = n$, the first two conditions coincide while the third is vacuous. We shall see later that for $m > n$ (2.26c) is a very strong condition on ρ.

We have demonstrated thus far that Eqs. (2.26) are necessary consequences of Eq. (2.15); if the latter is valid, there

is some mode function $u(x)$ with associated operator a, and constants C_m, C_n, such that Eqs. (2.26) are obeyed by ρ. The converse is also true: From Eqs. (2.26) we can obtain Eq. (2.15). We easily get the results, given (2.26):

$$\Gamma^{(m,n)}(x;y) = u(x_1)^*\cdots u(y_n)\ \mathrm{Tr}(a^n\rho a^{\dagger m}),$$
$$\Gamma^{(m,m)}(x;x') = u(x_1)^*\cdots u(x'_m)\ \mathrm{Tr}(a^m\rho a^{\dagger m}),$$
$$\Gamma^{(n,n)}(y;y') = u(y_1)^*\cdots u(y'_n)\ \mathrm{Tr}(a^n\rho a^{\dagger n}); \tag{2.27}$$

and from here, remembering Eq. (2.26c), the unimodularity of $S^{(m,n)}$ follows. Moreover, the factorizability of all correlation function $\Gamma^{(m',n')}$ for $m',n' \geqslant n$ is an easy consequence of Eqs. (2.26) [16]:

$$\Gamma^{(m',n')}(x;y) = g^{(m',n')}\, u(x_1)^*\cdots u(y_{n'}),$$
$$g^{(m',n')} = \mathrm{Tr}(a^{n'}\rho a^{\dagger m'}), \quad m',n' \geqslant n. \tag{2.28}$$

In particular, if the optical discriminant $\Delta^{(1,1)}$ vanishes so that the two-point correlation function $\Gamma^{(1,1)}(x;y)$ factorizes, then *all* correlation functions, *including* the ones $\Gamma^{(m,0)}$ and $\Gamma^{(0,m)}$, will factorize in terms of just one mode function $u(x)$, and the statistical state may be said to be "mode pure." This is not necessarily a coherent state, but it will be so provided

$$\mathrm{Tr}(a^n\rho a^{\dagger m}) = z^n z^{*m} \tag{2.29}$$

for some complex z and all $m,n \geqslant 0$.

The second part of the analysis concerns the most general form ρ can have if Eq. (2.15) is to be satisfied. We know that we must find the most general solution to Eqs. (2.26); of these, Eq. (2.26a) is a consequence of Eq. (2.26b) if $m > n$, so we need solve just Eqs. (2.26b) and (2.26c), which we write again for definiteness:

$$V(y_1)\cdots V(y_n)\,\rho = u(y_1)\cdots u(y_n)\, a^n\rho, \tag{2.30a}$$
$$a^m\rho = (C_m/C_n)\, a^n\rho. \tag{2.30b}$$

The case $m = n$ [when (2.30b) is empty] is taken up first, the case $m > n$ later. To begin with, we recall the general form of ρ, given that it is Hermitian, nonnegative, and of unit trace, its eigenvalues ρ_j form a discrete set summing up to unity. If the corresponding eigenvectors are $|\psi_j\rangle$, we have:

$$\rho = \sum_j \rho_j |\psi_j\rangle\langle\psi_j|\,, \quad \rho_j > 0, \quad \sum_j \rho_j = 1. \tag{2.31}$$

There are, naturally, no terms here corresponding to zero being a possible eigenvalue of ρ; the $|\psi_j\rangle$ will be an orthonormal set which is in general not a complete one but can always be extended to a complete orthonormal set. Now it can be seen that even if the vectors $|\psi_j\rangle$ are not pairwise orthogonal, the expression in Eq. (2.31) yields an acceptable density matrix provided only that the numbers ρ_j remain positive, sum up to unity, and each $|\psi_j\rangle$ is of norm unity. In this more general situation, we do not interpret the ρ_j and $|\psi_j\rangle$ as eigenvalues and eigenvectors of ρ, but interpret ρ as being a convex combination of the pure state density operators $|\psi_j\rangle\langle\psi_j|$. It follows that we need find the most general solutions to Eqs. (2.30) assuming ρ to be as given in Eq. (2.31) but need not insist that the $|\psi_j\rangle$ be pairwise orthogonal. But the fact that the constants ρ_j in Eqs. (2.30) are positive is enough to show that the operator conditions developed above for ρ must in fact be satisfied by each $|\psi_j\rangle$; it is not necessary, in order to make this statement, that (2.30) be an expansion of ρ in terms of its eigenvalues and eigenvectors. Specifically, each of Eqs. (2.16), (2.20), (2.24), (2.26), and (2.30) must remain valid if $|\psi_j\rangle$ replaces ρ in them.

For the case $m = n$, we choose any normalized mode function u and ask for the most general vector $|\psi_j\rangle$ obeying

$$V(x_1)\cdots V(x_m)|\psi\rangle = u(x_1)\cdots u(x_m)\, a^m|\psi\rangle. \tag{2.32}$$

If $u(x)$ is chosen as the first member of a complete orthonormal set $(u,v_1,v_2,\cdots)$ and the field $V(x)$ is expanded as

$$V(x) = u(x)\, a + \sum_\alpha v_\alpha(x)\, b_\alpha, \tag{2.33}$$

then the content of Eq. (2.32) can be stated in words thus: $|\psi\rangle$ must be annihilated by every product of m annihilation operators provided at least one of them is one of the b_α. $|\psi\rangle$ must be of the form

$$|\psi\rangle = \{F(a^\dagger) + P^{(m-1)}[V^\dagger(x)]\}|0\rangle, \tag{2.34}$$

where F is an arbitrary function of its argument and $P^{(m-1)}$ is a polynomial functional of the creation field $V^\dagger$ of degree not more than $(m-1)$. We may take any number of such (normalized) vectors $|\psi_j\rangle$, choose any positive constants ρ_j summing to unity, and put them into Eq. (2.31) to get the most general statistical state ρ for which the (m,m) optical discriminant vanishes identically. The quantities to be chosen freely are the mode function u, the functions F_j, the functionals $P_j^{(m-1)}$, and the ρ_j.

For the case $m = n + N$, $N \geqslant 1$, we choose again a mode $u(x)$ and a complex number z,

$$C_m/C_n = z^N. \tag{2.35}$$

We then look for the general solution $|\psi\rangle$ to

$$V(x_1)\cdots V(x_n)|\psi\rangle = u(x_1)\cdots u(x_n)\, a^n|\psi\rangle, \tag{2.36a}$$
$$a^N(a^n|\psi\rangle) = z^N(a^n|\psi\rangle). \tag{2.36b}$$

As in Eq. (2.34), Eq. (2.36a) is solved by

$$|\psi\rangle = \{F(a^\dagger) + P^{(n-1)}[V(x)^\dagger]\}|0\rangle, \tag{2.37}$$

but now the function F is severely restricted by the remaining condition (2.36b). In fact, apart from the freedom to add a polynomial of degree $(n-1)$ in $a^\dagger$ (which could be absorbed in $P^{(n-1)}$), we find

$$F(a^\dagger) = \sum_{r=0}^{N-1} \beta_r \exp(e^{2\pi i r/N} z a^\dagger), \tag{2.38}$$

where β_r are arbitrary constants. Thus the acceptable states $|\psi\rangle$ are largely determined in terms of coherent states as defined in the last section, with respect to the mode u:

$$|\psi\rangle = \sum_{r=0}^{N-1} \beta'_r |z e^{2\pi i r/N}\rangle_u + P^{(n-1)}[V(x)]^\dagger|0\rangle. \tag{2.39}$$

Once again taking several such (normalized) vectors $|\psi_j\rangle$ and choosing ρ_j, having first picked a $u(x)$ and a z, we get via Eq. (2.31) the most general statistical state with unimodular coherence index of order (m,n).

3. ENUMERATION OF EXCITED MODES

In the previous section we discovered the general form

of a statistical state for which it was assumed that some definite optical discriminant $\Delta^{(m,n)}$ vanished identically. The parts of ρ corresponding to "large" (greater than or equal to n) total photon numbers was to a considerable extent determined, and was seen to depend on just one single mode function $u(x)$. Now we outline a method of directly analyzing these components of ρ without assuming anything about the optical discriminants. As is to be expected, the large photon number description of ρ will involve more than just one mode function in general.

The total photon number operator, N, is given in terms of the potential $\mathbf{A}$ and field $\mathbf{E}$ as

$$N = -2i\int d^3r \mathbf{A}^{(+)}(\mathbf{r})^\dagger \cdot \mathbf{E}^{(+)}(\mathbf{r}). \tag{3.1}$$

Since $\mathbf{E}^{(+)}$ is $i\omega$ times $\mathbf{A}^{(+)}$, where ω is the frequency operator, we can rewrite N in the form

$$N = \int dx\, dy\, V(x)^\dagger H(x,y)\, V(y), \tag{3.2}$$

where we have denoted the Fourier transform of 2ω by $H(x,y)$, and the integrations include polarization sums. Let us now start with the (n,n) order correlation function $\Gamma^{(n,n)}(x;y)$,

$$\Gamma^{(n,n)}(x;y) = \mathrm{Tr}[V(y_1)\cdots V(y_n)\rho V(x_1)^\dagger \cdots V(x_n)^\dagger]. \tag{3.3}$$

Since the trace is unchanged by cyclic permutation of its arguments, and since any two V's ($V^\dagger$'s) commute, we can use Eq. (3.2) to "contract" one V and one $V^\dagger$ and produce an N at the proper place within the trace:

$$\begin{aligned}&\int H(x_1,y_1)\,\Gamma^{(n,n)}(x;y)\,dx_1 dy_1\\ &= \mathrm{Tr}[\rho V(x_2)^\dagger \cdots V(x_n)^\dagger N V(y_2)\cdots V(y_n)]\\ &= \mathrm{Tr}[V(y_2)\cdots V(y_n)(N-n+1)\rho V(x_2)^\dagger \cdots V(x_n)^\dagger].\end{aligned} \tag{3.4}$$

Carrying out this operation $(n-2)$ times more, and using cyclic invariance, we find:

$$\begin{aligned}&\int H(x_1,y_1)\cdots H(x_{n-1},y_{n-1})\,\Gamma^{(n,n)}(x;y)\,dz_1 dy_1 \cdots dx_{n-1}\,dy_{n-1}\\ &\quad = \mathrm{Tr}[V(y_n)(N-1)(N-2)\cdots(N-n+1)\rho V(x_n)^\dagger]\\ &\quad = \mathrm{Tr}[V(y_n)\sigma V(x_n)^\dagger],\\ &\sigma = \sqrt{(N-1)(N-2)\cdots(N-n+1)}\\ &\qquad \times \rho \sqrt{(N-1)(N-2)\cdots(N-n+1)}.\end{aligned} \tag{3.5}$$

The operator σ, like ρ, is Hermitian and nonnegative; all states in ρ with at most $(n-1)$ photons drop out in σ, while states in ρ with n or more photons survive with positive numerical factors. Thus by this process we obtain for $\Gamma^{(n,n)}$ an auxiliary two-point function,

$$\Phi(x,y) = \mathrm{Tr}[V(y)\,\sigma V(x)^\dagger], \tag{3.6}$$

which retains the positivity and hermiticity property but involves only n or more photon contributions to ρ. Let us then make an eigenvector decomposition of Φ:

$$\Phi(x,y) = \sum_\nu \lambda_\nu u_\nu(x)^* u_\nu(y), \quad \lambda_\nu > 0. \tag{3.7}$$

Only the strictly positive eigenvalues λ_ν of Φ and the corresponding eigenfunctions appear here; the functions u_ν form an orthonormal set which if not complete can be extended to a complete orthonormal set. Let u_ν, $\nu = 1,2,\cdots\infty$, be the latter complete set, and let the subset of ν values covered by the sum in Eq. (3.7) be denoted by κ. Make an expansion of V in the basis u_ν,

$$V(x) = \sum_{\nu=1}^{\infty} u_\nu(x) a_\nu. \tag{3.8}$$

Then Eq. (3.7) may be written

$$\sum_{\mu,\nu=1}^{\infty} u_\mu(x)^* u_\nu(y)\mathrm{Tr}(a_\nu \sigma a_\mu^\dagger) = \sum_{\nu\in\kappa} \lambda_\nu u_\nu(x)^* u_\nu(y), \tag{3.9}$$

that is,

$$\mathrm{Tr}(a_\nu \sigma a_\mu^\dagger) = \begin{cases} 0 & \text{if } \mu \neq \nu \text{ or } \nu \notin \kappa, \\ \lambda_\nu & \text{if } \mu = \nu \in \kappa. \end{cases} \tag{3.10}$$

Because of the basic nonnegativity of σ we may conclude from here that

$$a_\mu \sigma = 0 \quad \text{if } \mu \notin \kappa. \tag{3.11}$$

Moreover, since the most general structure for σ is of the form

$$\sigma = \sum_j \sigma_j |\phi_j\rangle\langle\phi_j|, \quad \sigma_j > 0, \tag{3.12}$$

with the (normalized but not necessarily mutually orthogonal) states $|\phi_j\rangle$ having at least n photons, the condition (3.11) passes over to each $|\phi_j\rangle$,

$$a_\mu |\phi_j\rangle = 0 \quad \text{if } \mu \notin \kappa. \tag{3.13}$$

For any given n, this analysis tells us that only the eigenmodes corresponding to nonzero eigenvalues of the auxiliary two-point function Φ can be excited in the contributions to ρ having at least n photons.[16] If u_ν, $\nu \in \kappa$, is less than a complete set, this is nontrivial information concerning ρ. The general solution to Eq. (3.13) is

$$|\phi\rangle = F^{(\geq n)}(a_\nu^\dagger, \nu\in\kappa)|0\rangle, \tag{3.14}$$

where $F^{(\geq n)}$ is a possibly infinite polynomial in the indicated operators with the lowest degree terms being at least of order n. Taking several ρ solutions of this form we can get the general structure of σ and then of ρ remembering Eq. (3.5):

$$\begin{aligned}&\rho = \sum_j \rho_j |\psi_j\rangle\langle\psi_j|, \quad \rho_j > 0, \quad \sum_j \rho_j = 1,,\\ &|\psi_j\rangle = \{F_j(a_\nu^\dagger, \nu\in\kappa) + P_j^{(n-1)}[V(x)^\dagger]\}|0\rangle.\end{aligned} \tag{3.15}$$

We assume the $|\psi_j\rangle$ are unit vectors. $P_j^{(n-1)}$ are polynomial functionals of degree no more than $(n-1)$ in $V(x)^\dagger$; and the F_j are arbitrary functions of the indicated operators. We note that this result is the generalization of the results of Section 2, with the single mode u being enlarged to the set u_ν, $\nu\in\kappa$. The correlation function $\Gamma^{(n,n)}$ in this case involves the distinguished modes u_ν, $\nu\in\kappa$:

$$\begin{aligned}&\Gamma^{(n,n)}(x;y) = \sum_{\mu,\nu\in\kappa} u_{\mu_1}(x_1)^* \cdots u_{\nu_n}(y_n)\, g^{(n,n)}(\mu;\nu),\\ &g^{(n,n)}(\mu;\nu) = \mathrm{Tr}(a_{\nu_1}\cdots a_{\nu_n}\rho a_{\mu_1}^\dagger \cdots a_{\mu_n}^\dagger).\end{aligned} \tag{3.16}$$

But it is immediately obvious that a similar restriction to the

set κ occurs in all $\Gamma^{(m',n')}$ for $m',n' \geqslant n$. It is as if we may ignore all modes except those present in Eq. (3.7) for computing the "sufficiently high" order correlation functions.

Additional information on the correlation functions beyond (n,n) allows repetition of this analysis. We now supply the auxiliary $\Phi(x,y)$ with an index n. Then the $\geqslant n$ photon parts of the statistical state ρ can involve only the eigenmodes of $\Phi_n(x,y)$ (corresponding to nonzero eigenvalues!). If we have a sequence of auxiliary functions $\Phi_{n_1}, \Phi_{n_2}, \cdots$ with $n_1 < n_2 < \cdots$, then we have a corresponding sequence of sets of eigenmodes $[u_\nu(x), \nu \in k_1]$, $[u_\nu(x), \nu \in \kappa_2], \cdots$ with the inclusion relations $\kappa_1 \supseteq \kappa_2 \supseteq \cdots$. The higher the order of the correlation function considered, the fewer the contributing modes!

4. SECOND HILBERT SPACE AND COHERENT STATES

Given a Hilbert space $\mathscr{H}$, we can induce a vector space structure with an inner product among the linear operators $A,B,\cdots$ on $\mathscr{H}$ by defining

$$(A,B) = \mathrm{Tr}(A^\dagger B). \tag{4.1}$$

So the operators of Hilbert–Schmidt class with finite values of $\mathrm{Tr}(A^\dagger A)$ constitute a second Hilbert space $\mathscr{K}$ based on $\mathscr{H}$. In particular we may choose $\mathscr{H}$ as the representation space of the (unbounded) operators $a,a^\dagger$ or q,p associated with a quantum system with one degree of freedom.

Let A be a general operator on $\mathscr{H}$. We define a set of four "superoperators" which are operators on $\mathscr{K}$ and whose effects on A considered as a vector in $\mathscr{K}$ are given by:

$$\mathscr{A}_1 A = (aA - Aa^\dagger)/\sqrt{2}, \quad \mathscr{A}_1^\dagger A = (a^\dagger A - Aa)/\sqrt{2},$$
$$\mathscr{A}_2 A = -i(aA + Aa^\dagger)/\sqrt{2}, \quad \mathscr{A}_2^\dagger = i(a^\dagger A + Aa)/\sqrt{2}. \tag{4.2}$$

These hermiticity relations are in accordance with the definition (4.1) of the scalar product. By virtue of their definitions it follows that

$$[\mathscr{A}_1, \mathscr{A}_1^\dagger] = [\mathscr{A}_2, \mathscr{A}_2^\dagger] = 1,$$
$$[\mathscr{A}_1, \mathscr{A}_2] = [\mathscr{A}_1, \mathscr{A}_2^\dagger] = 0. \tag{4.3}$$

Hence these operators correspond to a quantum system with *two* degrees of freedom.[8]

The simultaneous coherent states corresponding to it, and $\mathscr{A}_2$ are supplied by the outer products of the coherent states in $\mathscr{H}$:

$$\mathscr{A}_1 |z_1\rangle\langle z_2| = \frac{z_1 - z_2^*}{\sqrt{2}} |z_1\rangle\langle z_2|,$$
$$\mathscr{A}_2 |z_1\rangle\langle z_2| = -i\frac{z_1 + z_2^*}{\sqrt{2}} |z_1\rangle\langle z_2|. \tag{4.4}$$

In particular,

$$\mathscr{A}_1 |z\rangle\langle z| = i\sqrt{2}\,\mathrm{Im}\, z |z\rangle\langle z|,$$
$$\mathscr{A}_2 |z\rangle\langle z| = -i\sqrt{2}\,\mathrm{Re}\, z |z\rangle\langle z|. \tag{4.5}$$

Hence from among all the coherent states $|z_1\rangle\langle z_2|$ in $\mathscr{K}$, those with $z_1 = z_2 = z$ correspond to pure imaginary eigenvalues for both $\mathscr{A}_1$ and $\mathscr{A}_2$. By a straightforward generalization of Eq. (1.20) to two degrees of freedom, we may now expand a density operator ρ on $\mathscr{H}$ considered as a vector in $\mathscr{K}$ in terms of the imaginary eigenvalue coherent states,

$$\rho = (1/\pi)\int \phi(z) |z\rangle\langle z| d^2 z. \tag{4.6}$$

$\phi(z)$ is the (distribution) weight function analogous to $v(r)$ in Eq. (1.20). To determine $\phi(z)$ by an equation similar to (1.22) [or (1.23)], we must get the description of $\mathscr{K}$ by Schrödinger wavefunctions like $\psi(q')$. We define

$$Q_1 = (\mathscr{A}_1 + \mathscr{A}_1^\dagger)/\sqrt{2}, \quad P_1 = -i(\mathscr{A}_1 - \mathscr{A}_2^\dagger)/\sqrt{2},$$
$$Q_2 = (\mathscr{A}_2 + \mathscr{A}_2^\dagger)/\sqrt{2}, \quad P_2 = -i(\mathscr{A}_2 - \mathscr{A}_2^\dagger)/\sqrt{2}. \tag{4.7}$$

On a general $A \in \mathscr{K}$ these act as

$$Q_1 A = \frac{1}{\sqrt{2}}[q,A], \quad P_1 A = \frac{1}{\sqrt{2}}\{p,A\},$$
$$Q_2 A = \frac{1}{\sqrt{2}}[p,A], \quad P_2 A = \frac{-1}{\sqrt{2}}\{q,A\}, \tag{4.8}$$

and the only nonvanishing commutators among them are

$$[Q_1,P_1] = [Q_2,P_2] = i. \tag{4.9}$$

The (generalized) "eigenstates" of Q_1 and Q_2 turn out to be the Weyl family of unitary operators $W(\alpha)$ defined in Eq. (1.14). If α is written as $r + is$, we find:

$$Q_1 W(r+is) = rW(r+is), \quad Q_2 W(r+is) = sW(r+is),$$
$$P_1 W(r+is) = \frac{i\partial}{\partial r} D(r+is),$$
$$P_2 W(r+is) = \frac{i\partial}{\partial s} W(r+is),$$
$$(W(r'+is'), W(r+is)) = \pi\delta(r'-r)\delta(s'-s). \tag{4.10}$$

These operators have general coherent state matrix elements which may be viewed as a generalization of Eq. (1.10),

$$\langle z'|W(\alpha)|z\rangle = [|z'\rangle\langle z|, W(\alpha)] = \exp(-\tfrac{1}{2}|\alpha|^2 - \tfrac{1}{2}|z|^2$$
$$-\tfrac{1}{2}|z'|^2 + \alpha z'^* - \alpha^* z + zz'^*). \tag{4.11}$$

With the operators $W(\alpha)$ forming an "ideal" basis for $\mathscr{K}$, any $A \in \mathscr{K}$ and in particular a density operator ρ has an expansion

$$\rho = \int t(\alpha)\, W(\alpha) d^2\alpha \tag{4.12}$$

with the "Weyl weight" playing the role of $\psi(q')$ in Sec. 1. The distribution $\phi(z)$ obeys the integral equations[18]

$$\int \phi(z) \exp(-|z-z'|^2) d^2 z$$
$$= \pi \int t(\alpha) \exp(-\tfrac{1}{2}|\alpha|^2 + \alpha z'^* - \alpha^* z')\, d^2\alpha, \tag{4.13}$$

which generalizes Eq. (1.21) and arises by taking the diagonal coherent state matrix element in Eqs. (4.6) and (4.12) and using Eqs. (1.11) and (1.15). The solution

$$\phi(z) = \int t(\alpha) \exp(\tfrac{1}{2}|\alpha|^2) \exp(\alpha z^* - \alpha^* z)\, d^2\alpha \tag{4.14}$$

is the analog of Eq. (1.22). We can also make use of the properties of coherent states to derive the known result[18]

$$\alpha(z) = \frac{\exp(|z|^2)}{\pi} \int \langle -z'|\rho|z'\rangle \exp(|z'|^2 + zz'^* - z^*z')\, d^2z', \tag{4.15}$$

which is in the pattern of Eq. (1.23).

5. DISCRETE DIAGONAL REPRESENTATIONS

The preceding section has shown that the problem of the diagonal coherent state representation (4.6) for a density operator is in essence the same as the problem of expanding a vector $|\psi\rangle$ in terms of a subset of "pure imaginary" coherent states as attempted in Eq. (1.20). But for vectors we know that even more economical subsets exist through which arbitrarily accurate expansions can be made. We can thus combine the properties of these characteristic sets recounted in Sec. 1 with the approach of Sec. 4 to get new diagonal approximations for density operators.

We begin with the one-degree-of-freedom case. Since by Eq. (4.5) "diagonal" outer products of coherent states correspond to pure imaginary eigenvalues for $\mathscr{A}_1$ and $\mathscr{A}_2$, we proceed as follows. We pick two discrete characteristic sets $\{i\sqrt{2}y_n\}$ and $\{-i\sqrt{2}\,x_m\}$, both consisting of points on the imaginary axis, so that x_m and y_n are real. Then the double sequence of points $\{i\sqrt{2}\,y_n, -i\sqrt{2}\,x_m\}$ is a characteristic set[11] in the product of the complex plane by itself. Consequently the simultaneous coherent states of $\mathscr{A}_1$ and $\mathscr{A}_2$ with respective eigenvalues $i\sqrt{2}\,y_n$, $-i\sqrt{2}\,x_m$ give a set of elements in $\mathscr{K}$ through linear combinations of which one can approximate any member of $\mathscr{K}$ arbitrarily closely. But by Eq. (4.5) these elements of $\mathscr{K}$ are just the elements

$$|z_{mn}\rangle\langle z_{mn}|, \quad z_{mn} = x_m + iy_n. \tag{5.1}$$

Consequently we can approximate any density operator arbitrarily closely (in Hilbert–Schmidt norm!) by a discrete sum

$$\sum_{m,n} \phi_{mn} |z_{mn}\rangle\langle z_{mn}|. \tag{5.2}$$

The number of terms needed and the coefficients ϕ_{mn} to be used both depend on the desired accuracy of approximation.[19] In the case with one degree of freedom and one operator a, the operator (5.2) describes an ensemble over the preassigned set of values $\{z_{mn}\}$ for a, with the real coefficient ϕ_{mn} being the quasiprobability associated with the realization z_{mn} for a. Provided the trace property is maintained in these approximations to ρ, the ϕ_{mn} must add up to unity but in general the nonnegativity of ϕ_{mn} cannot be guaranteed; thus Eq. (5.2) is *not* in general, a convex combination of the pure state density operators $|z_{mn}\rangle\langle z_{mn}|$.

Since a set of points with a finite limit point does form a characteristic set,[11] we can have approximations of the above type with all but a finite number of points z_{mn} inside one phase cell! In certain respects this would mimic an amplitude-stabilized (mode-pure) classical light beam, but *every* ρ could be reached this way. The quantum aspects are hidden in the rapid variations of ϕ_{mn} as z_{mn} varies inside the chosen phase cell and also as the degree of approximation is improved.

At the other extreme, we can try to choose the set $\{z_{mn}\}$ as uniformly spread out as possible over the complex plane: the basic limitation now comes from the condition (1.25) for a characteristic set if finite limit points are to be avoided. This condition is to be applied both to $\{x_m\}$ and $\{y_n\}$. Let us then demand that

$$\sum_{n=1}^{\infty} |x_n|^{-2-\epsilon_1} = \infty, \quad \sum_{n=1}^{\infty} |y_n|^{-2-\epsilon_2} = \infty, \tag{5.3}$$

for some positive ϵ_1 and ϵ_2. It is clear that x_n and y_n cannot depend linearly on n, since these series would then converge for any $\epsilon_1, \epsilon_2 \geqslant 0$. Among fractional power dependences, possible candidates are, for example,

$$y_n, x_n \sim n^{\alpha}, \quad \alpha < \tfrac{1}{2}. \tag{5.4}$$

This means that the number of points of the set $\{z_{mn}\}$ contained in a square of side L is approximately $L^{2/\alpha}$ which is, for large L, larger than L^2. In this sense, one needs much more than one point for phase cell to have a set of pure state density operators $|z_{mn}\rangle\langle z_{mn}|$ built from coherent states, with combinations of which any ρ can be approximated.

Formally, representations of the type (5.2) can be extended from one degree of freedom to the entire field. The operator $V(x)$ is expanded in some complete orthonormal set $u_\alpha(x)$,

$$V(x) = \sum_{\alpha} u_\alpha(x) a_\alpha. \tag{5.5}$$

For each mode α, we choose some mesh of points $z^{(\alpha)}_{mn}$ in the complex plane in the manner described for one degree of freedom. Then a density operator ρ for the whole field can be approximated through expressions of the form

$$\sum_{\{m,n\}} \phi(\{m_\alpha, n_\alpha\}) |\{z^{(\alpha)}_{m_\alpha n_\alpha}\}\rangle\langle\{z^{(\alpha)}_{m_\alpha n_\alpha}\}| \tag{5.6}$$

with

$$a_\beta |z^{(\alpha)}_{m_\alpha n_\alpha}\rangle = \{z^{(\beta)}_{m_\beta n_\beta}\} |\{z^{(\alpha)}_{m_\alpha n_\alpha}\}\rangle. \tag{5.7}$$

We interpret (5.6) as an ensemble over the preassigned set of realizations of $V(x)$ given by

$$V(x;\{m,n\}) = \sum_{\alpha} \{z^{(\alpha)}_{m_\alpha n_\alpha}\}\, u_\alpha(x). \tag{5.8}$$

This is of course a c-number function of x. As in the one degree of freedom case, the quasiprobability $\phi(\{m,n\})$ that $V(x;\{m,n\})$ is realized is real but may be negative. Even the term quasiprobability is used only figuratively since two coherent states are never orthogonal!

6. CONCLUDING REMARKS

In this paper we have studied the properties of statistical states describing general light beams from two points of view. On the one hand we have shown how from the analysis of coherence functions one can systematically look for simplicity in the large photon number sectors of a given state.

On the other hand, we have given new understanding of the diagonal coherent state representation, and in the process discovered new representations embodying very subtle features of quantum theory.

The structure of a state for which the coherence index of order (m,n) is unimodular is essentially determined, for photon numbers greater than or equal to n, by a single mode function and a sequence of constants. The striking differences between the cases $m = n$ and $m > n$ are worth pointing out. In the former, for instance, even if one knew that m photon states were present in ρ, nothing definite could be said about $m + 1$ and higher photon states; they may or may not be present in ρ. If $m > n$, on the other hand, the presence of m photon states in ρ guarantees the presence of states with arbitrarily large photon numbers, all in the mode $u(x)$ of course. This is because the relevant parts of ρ are determined in terms of coherent states which are superpositions of states will all possible numbers of photons.

Turning to the theory of discrete diagonal coherent state approximations to ρ, we would like to make two comments. The first is to clarify the situation concerning the use of a characteristic set of coherent states distributed "as uniformly as possible" over the phase plane. At the level of making approximations to *vectors* through linear combinations of coherent states, the example due to von Neumann shows that by taking one coherent state per unit phase cell we get a characteristic set. In other words, the set of points

$$z_{l,m} = \sqrt{\pi}\,(l + im), \quad l,m = 0, \pm 1, \pm 2, \cdots$$

is a characteristic set in the complex plane, and any $|\psi\rangle$ can be approximated through combinations of coherent states $|z_{l,m}\rangle$. But the problem of the discrete diagonal approximations to operators through forming linear combinations of

$$|z_{mn}\rangle\langle z_{mn}|$$

is *different*. Here, the real and imaginary parts, x_m and y_n, of Z_{mn} must be such that the set (x_m, y_n), essentially, must be characteristic in the product of the complex plane by itself. Equivalently, $\{x_m\}$ and $\{y_n\}$ must each be a real characteristic set, and this precludes $\{z_{mn}\}$ being distributed uniformly over the complex plane!

The second comment concerns the fact that one must be content with arbitrarily close approximations to ρ but may "never quite get there." To us this seems to point to the very thin line dividing "definition" and "existence." One can well imagine being completely innocent of the theory of distributions as yet conceiving of the possibility of the diagonal representation in its conventional form, Eq. (4.6). Faced with equations such as (4.13), (4.14), and (4.15) to find the weight function ϕ, one would be forced to say that in general no ϕ exists obeying these equations, but that with "good" functions ϕ_α in the formula (4.6) one can produce density operators ρ_α that are "arbitrarily close" to any given ρ! Indeed, distributions may usefully be thought of in this way. It seems that the situation is qualitatively quite similar with the discrete representations based on characteristic sets.

An interesting practical problem is to find ways of computing the coefficients ϕ_{mn} in the discrete diagonal approximation (5.2) to a given ρ for a desired accuracy. One may try to get ϕ_{mn} from a knowledge of the Weyl weight $t(\alpha)$ of ρ, but the technique of Fourier transformation seems not useful in this context. This seems to be a genuinely difficult problem; we have only succeeded in establishing the *existence* of these representations.

ACKNOWLEDGMENT

This work was supported in part by the Department of Energy in contract with the University of Texas at Austin under contract No. EY-76-S-05-3992.

[1]Max Planck, Verh. D. Phys. Ges. **2**, 237 (1900).
[2]S.N. Bose, Z. Physik **26**, 178 (1924).
[3]G.K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge U.P., London and New York, 1963).
[4]M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1975).
[5]R.J. Glauber, Phys. Rev. Lett. **10**, 84 (1963); Phys. Rev. **130**, 2529 (1963).
[6]E.C.G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
[7]J.R. Klauder, J. Math. Phys. **4**, 1055, 1058 (1963); **5**, 177 (1963).
[8]N. Mukunda and E.C.G. Sudarshan, "New Light on the Optical Equivalence Theorem and a New Type of Discrete Diagonal Coherent State Representation," Pramāṇa **10**, 227–38 (1978).
[9]E. Schrödinger, Naturwissenschaften **14**, 644 (1927).
[10]I.E. Segal, Illinois J. Math. **6**, 500 (1962); S.S. Schweber, J. Math. Phys. **3**, 831 (1962).
[11]V. Bargman, Commun. Pure Appl. Math. **14**, 187 (1961).
[12]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1955); J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968); A.M. Perelomov, Theor. Math. Phys. (Russian) **6**, 213 (1971). Perelomov shows that the example given by von Neumann $z = \sqrt{\pi}\,(l + im)$, $l,m = 0, \pm 1, \pm 2, \cdots$ is overcomplete but only by *one* state and is therefore a characteristic set in the sense of (1.24) but *not* in the sense that it remains a characteristic set if any finite number of points are removed from it. We thank John R. Klauder for information on Perelomov's work.
[13]P.A.M. Dirac, Proc. R. Soc. London **114**, 710 (1927).
[14]L. Mandel, E.C.G. Sudarshan, and E. Wolf, Proc. Phys. Soc. London **84**, 435 (1964).
[15]D. Gabor, J. Inst. Elec. Eng. **93**, 429 (1946); E. Wolf, Proc. R. Soc. A **230**, 246 (1955).
[16]E.C.G. Sudarshan, J. Math. Phys. Sci. **3**, 121 (1969); see also J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968), pp. 166–7.
[17]U.M. Titulaer and R.J. Glauber, Phys. Rev. **140**, B676 (1965).
[18]C.L. Mehta and E.C.G. Sudarshan, Phys. Rev. **138**, B274 (1965).
[19]R.E. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (Am. Math. Soc., Providence, 1934).
[20]E.P. Wigner, Phys. Rev. **40**, 749 (1932); J.E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1948); E.C.G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).

JMP0500I JMAPA,00MAY9790508

Theory of temporal pump stochasticity in stimulated Raman scattering in dispersionless media[a)]

D. Eimerl

Lawrence Livermore Laboratory, University of California, Livermore, California 94550
(Received 13 July 1978; revised manuscript received 29 December 1978)

A diagrammatic formalism is presented which gives exact results for the forward Stokes wave in dispersionless media in the field of a temporally stochastic pump. Both pump and Stokes waves are assumed to be plane waves, but transient effects are treated exactly. Results are derived in the Bourret approximation for the n-point functions ($n > 2$) but are exact to all orders in the mean pump field and its two-point function. The generalization to treat general n-point functions within this formalism is possible and the method of doing so is indicated. The partial summation of diagrams corresponding to previous treatments is identified and the next order terms are evaluated. It is found that for high gain systems, approximations made in previous work may be quite poor.

I. INTRODUCTION

One particularly important and interesting question in the stimulated scattering of light and other parametric processes concerns the effect of pump noise on the scattered waves. Recently there has been some interest in this; in particular the effect of pump noise on stimulated Raman scattering[1-15] has been studied both experimentally and theoretically. Among the motivations for this are the application to Raman lasers,[16] and the absence of monochromatic sources in the ultraviolet and x-ray regions.[8] Our motivation concerns the role of pump nonuniformity in excitation of instabilities in high intensity lasers, and the question of the reproducibility of the performance of a high intensity pulsed laser on a shot by shot basis.

The theory of nonmonochromatic pumping has been investigated extensively by D'yakov.[1-8] Using an analogy with light propagation in turbulent media, D'yakov[1] developed an equation of the Fokker–Planck type for the mean field and intensity in the first Stokes wave. The result is that the gain at the Stokes frequency is essentially zero until a critical pump intensity is approached; for intensities near this threshold the gain may be large and comparable to the gain under monochromatic conditions. The analysis applies to very broadband pumps, whose correlation function is a δ function, and dispersive media. The critical intensity is proportional to the dispersion and vanishes in nondispersive media. It is proportional to the bandwidth also. Thus the major result is that the gain is zero unless the gain under monochromatic pumping is sufficient to overcome the effects of dispersion. However, the Fokker–Planck equations give only the threshold; they do not give the gain above threshold. D'yakov[3,4] generalized the analsyis to include finite bandwidth pumps, in the Bourret approximation, where only the lowest order correlation function of the pump is included. The method uses an iterative procedure to solve the coupled differential equations of stimulated Raman scattering. The fields are written as a sum of a mean value and a fluctuating term, and the result of the iterative procedure is a perturbative expansion of the Stokes wave in the fluctuations of the pump. This expansion is truncated at lowest nontrivial order. The steady state gain coefficients thus obtained are always less than the gain coefficient for monochromatic pumping, but rapidly approach this limiting value as the critical intensity of the Fokker–Planck treatment is exceeded. The convergence of this procedure does not appear to have been studied. In particular, limits on the pump fluctuations required for convergence have not been given.

The theoretical work on the temporal pump fluctuations has been reviewed by Akhmanov *et al.*[6-8] The problem of spatial nonuniformity of the pump has been discussed by Pasmanik and Friedman.[9] Their technique is analogous to D'yakov's iterative solutions, and an analysis of the validity of the lowest order approximation is given.

Several experiments[8,10-15] to measure the gain have been performed. Akhmanov *et al.*[8] used a rhodamine-6G-dye laser to provide a broadband optical noise source and measured both forward and backward gains in liquid nitrogen. Zubarev *et al.*[10,11,13] used neodymium glass lasers exciting SRS in SF_6. The existence of a threshold was clearly demonstrated. Vakhonev[15] *et al.*, using an iodine photodissociation laser to excite SRS in liquid nitrogen, and Zubarev[14] have shown that spatial inhomogeneity can reduce the gain significantly. Definitive conclusions from the experimental results are difficult to obtain because the pumps used are incompletely characterized, and because the unfolding of gains requires a detailed spatial analysis. However, the existence of the threshold and its dependence on bandwidth appear to be verified.

Most of the work on noisy pumping appears to be con-

[a)]Work performed under the auspices of the U.S. Department of Energy under contract No. W-7405-Eng-48.

 0022-2488/79/081811-13$01.00

fined to the region near threshold. The extension of theory to a strongly overdriven instability where gains of the order of e^{20} or e^{30} are present is questionable in view of the convergence limitation. Experimentally only the threshold region has been investigated, using input Stokes pulses and measuring gains of less than e^3. Although the threshold region is important in the study of Raman lasers, it is not of interest in high intensity systems where the beam propagates through gases. From the point of view of stability and reproducibility of performance, it is more useful to study nondispersive media under conditions where the instability is strongly overdriven. The development of a formalism to study this region is needed, and it is the aim of this paper to develop it.

This paper studies temporal pump fluctuations in nondispersive media. The system studied here is one in which a plane-wave pump drives a plane-wave Stokes wave. The medium is assumed to be both loss-less and dispersionless, but the effects of slowly relaxing molecular vibrations are treated exactly. Since any depletion of the pump means that the instability has already become too strong, we shall confine ourselves to the region where pump depletion may be neglected. In the case of a monochromatic pump this system is one for which exact analytic solutions are available. The existence of this analytic solution allows a systematic treatment of pump stochasticity which gives exact results. In general, a noisy pump may be written

$$A_p = A + B, \tag{1.1}$$

where A is the mean field, a nonstochastic term, and B is the fluctuation, or noise. If B is set equal to zero, the analytic solution given by Wang[17] follows. The pump is a function of the single variable $\xi = t - z/c$, its running time, and the fluctuations satisfy

$$\langle B(\xi)B^*(\xi')\rangle = G(\xi,\xi'), \tag{1.2}$$

where G is a general function describing the pump. The higher order correlation functions are given by the Bourret approximation. For example

$$\langle B_1B_2B^*_3B^*_4\rangle = G_{13}G_{24} + G_{14}G_{23} \tag{1.3}$$

so that the pump has only lowest order correlations. Higher order correlations may be included using the methods developed in this paper, but results are given only in the Bourret approximation (1.3).

The basic technique used is functional analysis,[18] which is applied to the Stokes wave, and leads to a diagrammatic expansion similar to the Dyson's equations used by D'yakov. The expansion is summed exactly, for general A and G, and the results presented in an integral analytic form. The results are illustrated by some simple applications, which contain much of the basic behavior. The system displays both absolute and convective instabilities. As the fluctuations are increased from zero, the instability remains convective, but increases in strength. The increase is initially linear, but becomes nonlinear as G increases. At a certain threshold the instability becomes absolute. That is, the Stokes wave $E_s(\xi,z)$ increases with ξ for z fixed and no steady state configuration is reached until the pump is depleted. The threshold is evaluated approximately, but it is found that at the point where the instability becomes absolute, the fluctuations in the Stokes intensity are so large that the experiment is essentially irreproducible. The limitations on the pump fluctuations required by the repeatability of the experiment are given in Sec. VII.

The previous analyses of D'yakov,[1-8] and Pasmanik and Friedman,[9] apply to both dispersive and nondispersive media. Their results correspond to a partial summation of diagrams discussed in Sec. VI, the chain approximation, and follow from the lowest order chain approximation. The next order terms are evaluated and are found to be nonlinear in the pump fluctuations. On the other hand, the analysis of Akhmanov[8] includes pure noise pumping where $A_p = 0$. This corresponds to the loop approximation, discussed in Sec. IV. The gain formulas are again nonlinear in the intensity, and the region where Akhmanov claims and Stokes wave becomes narrowband, even under broadband pumping, is investigated.

The paper is organized as follows. The stochastic theory of stimulated Raman scattering is described in the next section, and is applied in the following section to the Stokes field E_s. (See Sec. III.) The diagrammatic expansion is treated in some detail in Sec. III, leading to the main result, Eq. (3.28). Section IV derives the diagrammatic expansion of the Stokes intensity, culminating in the expression (4.27) for the major contribution to the mean intensity. The variance in the ensemble of experiments to which this theory is applied is simply

$$\delta E^2 = \langle E_sE_s^*\rangle - \langle E_s\rangle\langle E_s^*\rangle \tag{1.4}$$

and is a direct measure of the reproducibility of the experiment.

Applications are discussed in the remaining sections. The initial behavior of the system is treated in Sec. V, and several important cases are discussed. This section is useful because the mathematics is most tractable in this region. In Sec. VI, the case where the noise is small compared to the coherent term is treated. It is described by a sum over a particular set of diagrams, known as the chain approximation. In Sec. VII, the opposite case of pure noise ($A_p = 0$) is treated, and corresponds to a different partial sum of diagrams, the loop approximation. The material is discussed generally in Sec. VIII, which also contains some speculations on the inclusion of spatial inhomogeneities.

This paper is the first attempt at a rigorous or exact treatment of this problem. It contains some surprises, such as the appearance of an absolute instability in the limit of long correlation time, and the very low variance of the Stokes wave found in the limit of strong broadband pumping. It shows that the mathematical techniques used near threshold can give incorrect results in strongly driven systems. Testing these techniques by comparing them with exactly solvable systems emphasizes the need for better approximations in high gain systems with no exact analytic solutions.

II. STOCHASTIC THEORY OF SRS

In the approximation where pump depletion and

bleaching of the medium may be neglected, SRS is described by two coupled equations.

$$\left(\frac{\partial}{\partial z} + c^{-1}\frac{\partial}{\partial t}\right) E_s = \sigma_1 A_p Q^*, \tag{2.1}$$

$$\left(\frac{\partial}{\partial t} + \Gamma\right) Q = \sigma_2 A_p E_s^* + N. \tag{2.2}$$

Here E_s is the Stokes field, A_p the pump, and Q is the optical phonon, which is dispersionless. N is a stochastic force describing the intrinsic noise of the medium. The pump is of the form (1.1) and is a function of a single variable, $\xi = t - z/c$, the running time of the pump. A detailed mathematical treatment of these equations has been given by Wang.[17] Defining

$$\theta(\xi,\xi') = 2\left[\sigma_1\sigma_2 z \int_{\xi'}^{\xi} d\xi'' |A_p(\xi'')|^2\right]^{1/2}, \tag{2.3}$$

The solution for the Stokes field is

$$E_s(\xi,z) = E_s(\xi,0) + \sigma_1\sigma_2 A_p(\xi)\int_{-\infty}^{\xi} d\xi' \Phi(\xi,\xi') e^{-\Gamma(\xi-\xi')}, \tag{2.4}$$

where the integrand is

$$\Phi(\xi,\xi') = A_p^*(\xi') E_s(\xi',0) I_1(\theta(\xi,\xi'))2z/\theta(\xi,\xi') + \int_0^z dz' N^*(\xi',z') I_0(\theta(\xi,\xi')), \tag{2.5}$$

where the functions I_1 and I_0 are (diverging) Bessel functions of imaginary argument. The solution for the Stokes field (2.4), is a functional of the stochastic variables A_p and N. Thus to determine the mean field, etc., the expectation value of products of E_s with either E_s or E_s^* with respect to the random fields must be calculated. By expanding the Bessel functions, obtaining products of powers of the stochastic variables, this can in principle be carried out term by term. However, this is prohibitively complicated to carry out beyond the lowest orders of approximation. Fortunately, a simpler technique is available, that of functional analysis.[18]

The technique may be illustrated by considering a simple function of a single variable $y = f(x)$. Assuming that f has a power series expansion near the point x, then the value of f at neighboring points is given by the formula

$$f(x + a) = \exp(ad/dx) f(x). \tag{2.6}$$

If Z is a functional of the field $Y(x)$, then a similar formula applies,

$$Z\{Y(x) + A(x)\} = \exp\int dx' A(x')[\delta/\delta Y(x')] Z\{Y(x)\}. \tag{2.7}$$

This may be proved by generalizing (2.6) to the case of many variables and taking a continuum limit. The functional derivative in the exponent has the following definition,

$$[\delta/\delta Y(x)] Y(z) = \delta(x - z). \tag{2.8}$$

To apply (2.7) to the Stokes wave in SRS, it is sufficient to note that the Stokes wave is a functional of two complex variables, A_p, and N. The dependence on N is linear. Consequently it is trivial to compute, and will not be considered any further in detail. For a general function F, a double application of (2.7) gives

$$F\{A_p(\xi), A_p^*(\xi')\} = \exp\left[\int d\eta\, B(\eta)[\delta/\delta A(\eta)] + B^*(\eta)[\delta/\delta A^*(\eta)]\right] \cdot F\{A(\xi), A^*(\xi')\}, \tag{2.9}$$

where now all the stochasticity appears in the exponent operator. The expectation value of F is obtained by expanding the exponent, and resuming. The result is

$$\langle F\rangle = e^K F\{A, A^*\}, \tag{2.10}$$

where K is a differential operator,

$$K = \sum_{n,m}\int d\eta_1\cdots d\eta_n d\lambda_1\cdots d\lambda_m\, G^{n,m}(\eta_1\cdots\eta_n\lambda_1\cdots\lambda_m)(n!m!)^{-1} \times [\delta/\delta A(\eta_1)]\cdots[\delta/\delta A(\eta_n)] \times [\delta/\delta A^*(\lambda_1)]\cdots[\delta/\delta A^*(\lambda_m)]. \tag{2.11}$$

In (2.11) the correlation function is the irreducible part of the expection value of the product of n B-fields and m B^*-fields. In the Bourret approximation we use, only the lowest order term survives,

$$K_1 = \int d\eta\, d\lambda\, G(\eta,\lambda)[\delta/\delta A(\eta)]\,[\delta/\delta A^*(\lambda)]. \tag{2.12}$$

The advantage of this functional approach is that all the manipulations involving stochastic variables have been carried out, for a general function F. The cost of this closed form result is the appearance of the functional derivatives. In general (2.10) contains an infinite number of derivatives unless F has the form of a finite polynomial, a condition which is not satisfied by the Stokes field. However, the Stokes wave has the form of an infinite power series in θ, which is a simple functional of the pump. This allows a systematic treatment of the Stokes wave to all orders. The Stokes field may be written as the sum of three terms,

$$E_s(\xi, z) = E_s(\xi,0) + \int_{-\infty}^{\xi} d\xi' e^{-\Gamma(\xi-\xi')}(\phi_1 + \phi_2) E_s(\xi',0). \tag{2.13}$$

In this formula

$$\phi_1(\xi',z) = \sum_n \frac{(\sigma_1\sigma_2 z)^{n+1}}{(n+1)!}\frac{\chi^n}{n!} A^*(\xi') A(\xi) \tag{2.14}$$

and

$$\phi_2(\xi',z) = \sum_n \frac{\chi^n}{n!} A(\xi)\int_0^z dz' \frac{(\sigma_1\sigma_2 z')^n}{n!} N^*(\xi',z'), \tag{2.15}$$

where χ is a simple function of the pump,

$$\chi = \int_{\xi'}^{\xi} d\xi'' A(\xi'') A^*(\xi''). \tag{2.16}$$

Consequently, the expectation value of the Stokes field (or intensity) involves the operation of e^K on a very low order polynomial in A and A^* times arbitrarily high powers of χ.

Thus, the expectation value of the field is

$$\langle E_s(\xi,z)\rangle = e^K E_s\{A,A^*\}, \tag{2.17}$$

and expansion of the exponent, together with (2.13) shows that we require the general term $R_{n,m}$, where

$$R_{n,m} = (n!)^{-1}K^n(\chi^m(\xi,\xi')A(\xi)A^*(\xi')). \tag{2.18}$$

If the intensity is required, then we need the term $P_{n,m,m'}$, where

$$P_{n,m,m'} = (n!)^{-1}K^n(\chi^m(\xi\xi_1)\chi^{m'}(\xi,\xi_2)A(\xi)A^*(\xi)A(\xi_1)A^*(\xi_2)) \tag{2.19}$$

in order to compute $e^K|E_s|^2$. This is discussed more fully later.

III. Diagrammatic Analysis for the Stokes Field

The series (2.17) for the Stokes field requires the general term $R_{n,m}$. Since $R_{n+1,m} = KR_{n,m}$, there is a recursion relation, and we consider the first element $R_{1,m}$. By carrying out the indicated differentiations, we obtain five terms,

$$R_{1,m} = \int d\eta\, d\eta'\, G(\eta,\eta')$$

$$\times [m(m-1)\chi^{m-2}A(\eta)A^*(\eta')A(\xi)A^*(\xi') \tag{3.1}$$

$$+ m\chi^{m-1}\delta(\eta-\eta')A(\xi)A^*(\xi') \tag{3.2}$$

$$+ m\chi^{m-1}\delta(\eta'-\xi)A(\eta)A^*(\xi') \tag{3.3}$$

$$+ m\chi^{m-1}\delta(\eta-\xi')A(\xi)A^*(\eta') \tag{3.4}$$

$$+ \chi^m\delta(\eta'-\xi)\delta(\eta-\xi')]. \tag{3.5}$$

These terms may be represented diagrammatically for compactness. In Fig. 1 each term in $R_{1,6}$ is displayed. Each χ is represented by a pair of dots placed in a vertical line. The unmatched dots represent the fields at ξ and ξ'. The upper row of dots is associated with A, the lower row with A^*, and

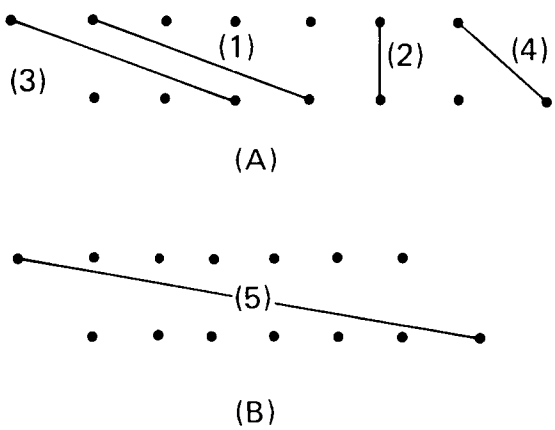

FIG. 1. Elements of the diagrammatic series for the Stokes field. Each pair of dots represents χ, each line a correlation function. These diagrams represent terms in $R_{1,6}$, but superposed for compactness.

each pair represents the fields at a point ξ_i, for the ith pair. Each line in the diagram represents the correlation function G, so that $R_{1,6}$ is obtained as the sum of all diagrams containing six pairs in which one line has been inserted. Thus, the line (1) represents the term (3.1), etc., and the diagrammatic series for $R_{1,6}$ contains 49 terms. There are thirty diagrams of type (1), six each of type (2), (3), and (4), and one of type (5). (Most of these have been superposed on one diagram in Fig. 1 to save space.) The rules by which algebraic expressions are associated with diagrams are straightforwardly obtained from $R_{1,m}$. They are as follows:

(1) Each free dot represents $A(\xi_i)$ $(A^*(\xi_i))$ if it is in the upper (lower) row;

(2) Each attached dot represents unity;

(3) Each line joins a dot in the upper row at ξ_1 with one in the lower row at ξ_2, and represents $G(\xi_1,\xi_2)$;

(4) Each intermediate coordinate is integrated from ξ to ξ', the coordinates of the unpaired dots.

Using the recursive relation for $R_{n,m}$, it follows that it possesses a diagrammatic expansion exactly analogous to $R_{1,m}$. Thus, $R_{n,m}$ is represented by the set of diagrams in which n lines are inserted simultaneously, and interpreting it according to the rules just given. This is most simply demonstrated by computing $R_{2,m}$ and then proving that each application of K_1, the lowest order term in K, replaces one A and A^* by a correlation function. The permutation combinations may easily be shown to be given correctly by this procedure. In this way, the complicated algebraic problem is reduced to a simpler topological one.

Consider the general term $R_{n,m}$. It is the sum of the n-line, m-pair diagrams. Each diagram in this series may be subdivided into simple irreducible components. A diagram or a subdiagram is termed irreducible if it cannot be separated into two parts without severing at least one line. The separation test refers to separating pairs from each other and not to separating dots from the same pair. Then a loop is defined as an irreducible diagram containing as many lines as pairs. Thus a loop cannot involve one of the unpaired dots, and has no unattached dots. A chain is defined as an irreducible diagram containing exactly two unattached dots. Thus, a chain cannot involve one of the unpaired dots either. However, the connected part of the diagram is defined as all irreducible parts which involve the unpaired dots. These have the same topological structure as chains, extending into the diagram from the unpaired dots at each end. Thus, the connected part is either the product of two chainlike terms, or consists of a single chainlike term involving both unpaired dots. In this way an arbitrary diagram may be written in terms of its irreducible components: the connected part, chains, loops, and free pairs. Intuitively, a loop is a ring of pairs, a chain is a chain of pairs and the connected part contains the chains of pairs which are connected to the unpaired dots. The components are illustrated in Fig. 2.

This decomposition allows the full series for the Stokes field to be written as an integral over sums and products of chains and loops. For example, the 3-loop is

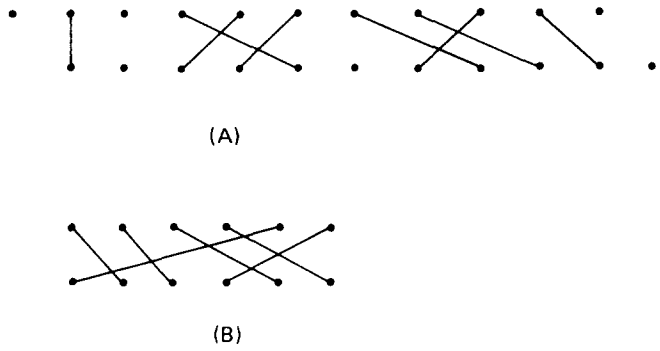

FIG. 2. (A) A term in $R_{8,10}$ containing a 4-chain, a 3-loop, a 1-loop, and a free pair. (B) A reducible decomposable into a 4-loop and a 2-loop.

$$L_3 = \int_{\xi'}^{\xi} d\xi_1 d\xi_2 d\xi_3\, G(\xi_1,\xi_2) G(\xi_2,\xi_3) G(\xi_3,\xi_1), \tag{3.6}$$

and the 2-chain is

$$C_2 = \int_{\xi'}^{\xi} d\xi_1 d\xi_2\, A^*(\xi_1) G(\xi_1,\xi_2) A(\xi_2). \tag{3.7}$$

If the B field is Gaussian noise and the A field is slowly varying, these are simple integrals obtainable analytically in terms of tabulated functions.

The general term $R_{n,m}$ may be decomposed into a connected part and a remaining factor, the vacuum diagrams.[19] Thus,

$$R_{n,m} = \sum_{pq} V_q^p H_{m,q}^{n-p}, \tag{3.8}$$

where V denotes the vacuum diagrams involving p lines and q pairs, and $H_{m,q}^{n-p}$ denotes the connected piece in which $n-p$ lines are inserted into m pairs leaving exactly q pairs untouched. The number of ways this can be done is given by the binomial coefficients. Thus if h_c^b denotes the connected pieces involving b lines and c pairs exactly, leaving no untouched pairs, then

$$R_{n,m} = \sum_{pq} V_q^p h_{m-q}^{n-p} C_q^m, \tag{3.9}$$

where C_q^m is the binomial coefficient $m!/q!(m-q)!$. The vacuum diagrams are therefore those diagrams in which p lines are inserted into q pairs without involving an unpaired or external dot. In general V involves free pairs. Their contribution to V may be separated out as follows,

$$V_q^p = \sum_l \chi^l U_{q-l}^{p-l} C_l^q. \tag{3.10}$$

Each free pair contributes χ, and the remaining part of the vacuum diagram, U_c^b is defined as the set of diagrams in which b lines are inserted into c pairs, with no free pairs left over and no unpaired dots involved. Thus U must be a sum over loops and chains. The lowest order terms in U are as follows:

$$U_0^0 = 1, \quad U_1^1 = L_1, \quad U_2^2 = L_2 + (L_1)^2, \tag{3.11}$$

$$U_1^1 = 2C_2. \tag{3.12}$$

The number of chains in U_s^r is given in general by $n_c = s - r$. If r lines are inserted into s pairs, there must be $2(s-r)$ unattached dots, and since each chain involves exactly two unattached dots, the result follows.

Separation of U into loops and chains is achieved by considering that part of U in which k lines and pairs form loops. Thus,

$$U_s^r = \sum_k U_L^k U_{cs-k}^{r-k} C_k^s. \tag{3.13}$$

From (3.13) and (3.10) it follows that V_q^p contains exactly $q-p$ chains.

The chain term U_c may be written in terms of the irreducible diagrams by dividing the k pairs into collections of k_j, associating each k_j with a loop and summing over all possible decompositions. Division of k pairs into collections of k_j can be accomplished in $k!/\Pi k_j!$ ways. Each group k_j forms a chain $2k_j!$ ways. Thus, if each chain is of different length, each element in the chain term U_c has the form

$$U_{cs}^r \sim 2^{n_c} s! C_{s_1} C_{s_2} \cdots C_{s_{n_c}}, \tag{3.14}$$

where $s = \Sigma_j^{n_c} s_j$. However, if two chains have the same length, then double counting must be avoided. The result for the chain term is as follows,

$$U_{cs}^r = s! \sum_{\{t_k\}} \prod_{k=2}^{\infty} \frac{(2C_k)^{t_k}}{t_k!} \delta\left[n_c, \sum t_k\right] \delta\left[s, \sum k t_k\right]. \tag{3.15}$$

For each chain of length k, the chain may appear in U_c, t_k times. All possible appearances and powers must be included, consistent with the constraints on the number of pairs and chains.

Similarly, the loop terms U_L may be written in terms of the irreducible loops. However, each group of k_j pairs may form an irreducible loop $(k_j - 1)!$ ways. To prove this, consider the loop with b pairs. The next pair may be included b ways to form the irreducible loop with $b+1$ pairs. Consequently each diagram for b pairs gives b diagrams for $b+1$ pairs. Induction then gives $(b-1)!$ for the degeneracy of the b-loop. Thus, the loop term is

$$U_L^b = b! \sum_{\{u_k\}} \prod_{k=1}^{\infty} \frac{(L_K/k)^{u_k}}{u_k!} \delta\left[b, \sum k u_k\right]. \tag{3.16}$$

The derivation of (3.16) is entirely analogous to that of (3.15).

The connected piece may be simplified by recalling that it has a particularly simple topological form. In fact, it is either the product of two chainlike terms or a single chainlike term,

$$h_s^r = s!\,[\delta_{r,s+1} h_s^{(1)} + \delta_{r,s} h_s^{(2)}] \tag{3.17}$$

In this expression

$$h_s^{(1)} = \int_{\xi'}^{\xi} d\xi_1 d\xi_2 \cdots d\xi_s\, G(\xi',\xi_1) G(\xi_1,\xi_2) \cdots G(\xi_s,\xi) \tag{3.18}$$

and

$$h_s^{(2)} = \sum_i h_i^{(3)} h_{s-i}^{(4)}, \tag{3.19}$$

where

$$h_i^{(3)} = \int_{\xi'}^{\xi} d\xi_1 d\xi_2 \cdots d\xi_i\, G(\xi',\xi_1) G(\xi_1\xi_2) \cdots G(\xi_{i-1},\xi_1) A(\xi_i) \tag{3.20}$$

and

$$h_i^{(4)} = \int_{\xi'}^{\xi} d\xi_1 d\xi_2 \cdots d\xi_i A^*(\xi_i) G(\xi_i,\xi_{i-1}) \cdots G(\xi_2,\xi_1) G(\xi_1,\xi'). \tag{3.21}$$

The essential differences between these expressions and the chain terms are apparent. Note in (3.18) that there are $s+1$ correlation functions in $h_s^{(1)}$.

The remaining step in the diagrammatic analysis is to collect the various terms in the Stokes field expansion (2.13). Applying the noise operator to each term in ϕ_1 [Eq. (2.14)], gives

$$\begin{aligned} e^K (A(\xi) A^*(\xi') \chi^n / n!) &= \sum_r R_{r,n}/n! \\ &= \sum_{q=0}^{n} R_q (h_{m-q}^{(1)} + h_{m-q}^{(2)}), \end{aligned} \tag{3.22}$$

where (3.9) and (3.17) have been used. The term R_q is the vacuum contribution,

$$R_q = \sum_p V_q^p / q!. \tag{3.23}$$

Using the previous expressions for V, this may be expressed in the following form,

$$R_q = \sum_{plkr} \frac{\chi^l U_L^k U_{cr}^p}{l!k!r!} \delta[q,r+k+l]. \tag{3.24}$$

The δ function may be written as a contour integral

$$\delta[q,r+k+l] = (2\pi i)^{-1} \int d\omega\, \omega^{r+k+l-q}, \tag{3.25}$$

where the contour is any region including the origin. Then the vacuum term R_q separates into three terms, each of which separates into simple components. The result is a simple integral expression

$$R_q = \int \frac{d\omega\, e^P}{2\pi i \omega^{1+q}} \tag{3.26}$$

where the exponent P is a sum over the chains and loops,

$$P(\omega) = (\chi + L_1)\omega + \sum_{k=2}^{\infty} (2C_k + L_k/k)\omega^k \tag{3.27}$$

Finally, the expression for ϕ_1 in (2.14) is obtained,

$$e^K \phi_1 = \int \frac{d\omega e^P}{2\pi i \omega} f_1(\omega), \tag{3.28}$$

where

$$f_1(\omega) = \sum_{n,q} \frac{(\sigma_1\sigma_2 z)^{n+q+k}}{(n+q+1)!} (h_n^{(1)} + h_n^{(2)}) \omega^{-q}. \tag{3.29}$$

Note that P and f_1 are functions of ξ and ξ', and that ω has the dimensions χ^{-1}, as written. Appropriate dimensional parameters may be inserted to make ω dimensionless, if it proves useful to do so.

IV. DIAGRAMMATIC ANALYSIS FOR THE STOKES INTENSITY

The exact result for the Stokes intensity is obtained from (2.4). However, rather than computing the Stokes intensity directly, it is more convenient to evaluate the noise at the Stokes frequency, by examining the mean square deviation of the field amplitude. This is a real positive quantity defined as follows,

$$\delta E^2 = \langle EE^* \rangle - \langle E \rangle \langle E \rangle^*. \tag{4.1}$$

Expanding the Bessel functions in (2.5), one obtains a form for the mean square field deviation similar to (2.13), etc., for the mean field. The major differences are the appearance of a term involving the intrinsic noise of the medium, and a double integral representation,

$$\begin{aligned} \delta E^2 = &\int_{-\infty}^{\xi} d\xi \exp[-\Gamma(\xi-\xi_1)] E_s(\xi_1,0) \\ &\times \int_{-\infty}^{\xi} d\xi_2 \exp[-\Gamma(\xi-\xi_2)] E_s^*(\xi_2,0)(\psi + \psi_N). \end{aligned} \tag{4.2}$$

In this expression the integrand has the following form,

$$\psi = e^K \phi_1(\xi_1,z)\phi_1^*(\xi_2,z) - e^K \phi_1(\xi_1,z) e^K \phi_1^*(\xi_2,z). \tag{4.3}$$

Here ϕ_1 is given by (2.14), and e^K is the noise operator introduced earlier. The second term in the integrand depends on the correlation function of the noise source. Because the velocity of the optical phonon may be neglected, this source has no correlation between neighboring points. Thus

$$\langle N(z,t) N^*(z',t') \rangle = N^2 \delta(z-z') G_N(\xi - \xi') \tag{4.4}$$

and its contribution to the integrand is

$$\psi_N = N^2 G_N(\xi_1 - \xi_2) e^K |A(\xi)|^2 \times \sum_{n,m} \frac{(\sigma_1\sigma_2 z)^{n+m+1} \chi(\xi,\xi_1)^n \chi(\xi,\xi_2)^m}{\sigma_1\sigma_2 (n+m+1)(n!m!)^2}. \tag{4.5}$$

[Akhmanov *et al.*[8] assume that G_N is δ-correlated. However, this implies that the relaxation time of the thermal noise in the optical phonon is much shorter than the transverse relaxation time of the molecular excitation. This appears to be questionable, since both phenomena originate in the same mechanism, namely dephasing collisions. Thus G_N has a correlation time comparable to $1/\Gamma$ and cannot be set equal to a δ function under the ξ integrals. Equation (4.5) is the basis for Akhmanov's treatment of the case of slowly relaxing molecular oscillations.]

In order to compute the variance, δE, and consequently to obtain the intensity, we shall, for the moment, ignore the term ψ_N, and note that the second term in (4.3) has already been obtained in the previous section. The first term, through (2.14), gives rise to expressions of the form of $P_{n,m,m'}$, defined in (2.19). These may be evaluated by a diagrammatic approach which is exactly analogous to that used for the field. An inessential complication is the appearance of three times, ξ, or equivalently the appearance of two types of pairs, depending on whether the arguments in χ are (ξ,ξ_1)

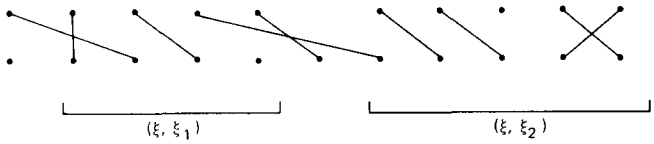

FIG. 3. Elements in the diagrammatic expansion for δE. The first pair is not integrated over, but is evaluated at ξ. The term displayed is a contribution to $P_{9,4,5}$.

or (ξ,ξ_2). In Fig. 3, a typical term in the diagrammatic expansion for the variance is displayed. The unmatched dots represent the field at ξ_1 and ξ_2, and the first pair (lhs) is not a pair to be integrated over, but represents the product $A(\xi)A^*(\xi)$. The pairs are divided into two groups, which may be connected by lines regardless of the group to which they belong. The rules for interpretation of the graphs are otherwise the same as for the mean field diagrams. The result is a decomposition of diagrams into connected and vacuum parts, exactly as above, and subsequent decomposition into chains and loops. However, these terms must be generalized somewhat in order to accommodate the existence of two types of pair, and the ordering of the integrations associated with a particular sequence of pairs.

The general term $P_{n,m,m'}$ may be decomposed into a connected part and a remaining factor, the vacuum diagrams. Thus

$$P_{n,m,m'} = \sum_{pqq'} V^{p}_{qq'} H^{n-p}_{mq\,m'q'}, \tag{4.6}$$

where V denotes the vacuum diagrams involving q pairs of type (1), (ξ,ξ_1) and q' of type (2). Clearly, the decomposition into simpler parts is exactly analogous to the mean field case, but where the indices must follow each type of pair explicitly. Thus, in the separation into vacuum and connected pieces, and the subsequent decomposition of the vacuum terms, the variance expansion carries two sets of indices, whereas the mean field expansion carries one set. This is the only major difference in the systematics of the expansions. However, the terms which can contribute are more varied and numerous in the variance expansion.

The systematics of the decomposition of the vacuum terms into chains and loops have already been detailed. On carrying out this process one obtains the following result,

$$\begin{aligned} R_{qq'} &= \sum_{p} V^{p}_{qq'}/q'q'! \\ &= (2\pi i)^{-2}\int d\omega_1 d\omega_2\, e^{Q}/\omega_1^{1+q}\omega_2^{1+q'}. \end{aligned} \tag{4.7}$$

The exponent R is a sum over generalized loops and chains

$$\begin{aligned} Q(\omega_1,\omega_2) &= \chi_1\omega_1 + \chi_2\omega_2 + \sum (2C_l + L_{kl}/(k+l)) \\ &\quad \times \frac{\omega_1^k}{k!}\frac{\omega_2^l}{l!}(k+l)!, \end{aligned} \tag{4.8}$$

where $\chi_{1,2}$ are given by

$$\chi_{1,2} = \int_{\xi_{1,2}}^{\xi} |A(\xi'')|^2 d\xi'', \tag{4.9}$$

and the loops and chains are as follows:

$$L_{kl} = k!l!/(k+l)! \sum_{\sigma} L^{\sigma}_{kl}, \tag{4.10}$$

$$C_{kl} = k!l!/(k+l)! \sum_{\sigma} C^{\sigma}_{kl}. \tag{4.11}$$

The parameter σ is a label taking $(k+l)!/k!l!$ values, which labels the ordering of the integrations in these elements. For example, L_{k0} is proportional to L_k, containing only one type of integration, whereas L^{σ}_{k1} contains k integrations over the range (ξ,ξ_1) and one over the range (ξ,ξ_2). Thus each permutation of the orders of the integration limits must be included. All the degeneracy factors have been removed in (4.8) and in the definitions (4.10) and (4.11) so that only distinct orderings are to be included in the sums indicated in L_{kl} and C_{kl}. Each ordering has weight unity. Finally the sum in (4.8) is over the range $0 \leqslant k,l < \infty$, but excluding the point $k = l = 0$.

For direct comparison with the mean field expansion, the 2-, 1-loop is

$$\begin{aligned} L_{21} &= \int_{\xi_1}^{\xi} d\xi' \int_{\xi_1}^{\xi} d\xi'' \int_{\xi_2}^{\xi} d\xi''\, G(\xi',\xi'') \\ &\quad \times G(\xi'',\xi''',)G(\xi''',\xi'). \end{aligned} \tag{4.12}$$

There is only one type of 2-, 1-loop, since all others are the same as this by permutation symmetry. There are, however, three distinct 2-, 1-chains, according to the position of the (ξ,ξ_2) integration.

In the case $\xi_1 = \xi_2$ the distinction disappears and at this point we have

$$L_{kl} = L_{k+l}, \quad C_{kl} = C_{k+l}, \tag{4.13}$$

and

$$Q(\omega_1,\omega_2) = P(\omega_1+\omega_2). \tag{4.14}$$

The distinction between field- and variance-expansion quantities is in the number of arguments or indices involved. It is straightforward to show that in this case, $(\xi_1 = \xi_2)$, one of the ω-integrals may be performed analytically, giving the following result, which is necessary for consistency,

$$R_{qq'} = \frac{(q+q')!}{q!q'!} R_{q+q'}. \tag{4.15}$$

The connected pieces form three categories according to the number of chainlike terms in them. From (4.6)

$$H^{p}_{mq,m'q'} = C^{m}_{q} C^{m'}_{q'} h^{p}_{m-q,m'-q'} \tag{4.16}$$

in a simple generalization of (3.9), and the reduced term is expressible as a sum over products of chainlike terms,

$$\begin{aligned} h^{r}_{st} &= s!t!\,[\delta[r,s+t+2]h^{(1)}_{st} + \delta[r,s+t+1]h^{(2)}_{st} \\ &\quad + \delta[r,s+t\,]h^{(3)}_{st}]. \end{aligned} \tag{4.17}$$

The first term is a product of two chainlike terms,

$$\begin{aligned} h^{(1)}_{st} &= \sum_{ij} (b_{s-1,t-j}(\xi,\xi_1)b_{ij}(\xi_2,\xi) \\ &\quad + b_{s-i,t-j}(\xi,\xi)b_{ij}(\xi_1,\xi_2))\, C^{s}_{i} C^{t}_{j}/s!t!, \end{aligned} \tag{4.18}$$

where each term is a sum over different orderings of the integration limits. For example,

$$b_{rs}(\xi,\xi_1) = \sum_{\sigma} b^{\sigma}_{rs}(\xi,\xi_1), \tag{4.19}$$

where

$$b^{\sigma}_{rs}(\xi,\xi_1) = \int d\xi_2 \cdots d\xi_\beta G(\xi,\xi_2)\cdots G(\xi_\beta,\xi_1) \tag{4.20}$$

contains r (ξ,ξ_1)-limits and s (ξ,ξ_2)-limits, and σ labels their order of appearance. Note that each ordering is distinct, so that there are C^{r+s}_{r} terms. The second term is a sum products of three chainlike terms,

$$h^{(2)}_{st} = P_{st}(\xi,\xi_1;\xi_2,\xi) + P_{st}(\xi,\xi;\xi_2,\xi_1) + P_{st}(\xi_2,\xi;\xi,\xi_1) \tag{4.21}$$

where

$$P_{st}(\xi,\xi_1;\xi_2,\xi') = \sum_{ijklmn} b_{il}(\xi,\xi_1) d^{*}_{jm}(\xi') d_{kn}(\xi_2) \times (i!j!k!l!m!n!)^{-1}\delta[s,i+j+k]\delta[t,l+m+n] \tag{4.22}$$

and d is a sum over integration orderings, $d = \Sigma d^{\sigma}$,

$$d^{\sigma}_{jm}(\xi) = \int d\xi_\alpha \cdots d\xi_\beta A(\xi_\alpha) G(\xi_\alpha,\xi_{\alpha+1})\cdots G(\xi_\beta,\xi). \tag{4.23}$$

Again all orderings are distinct. The last term is a product of four chainlike terms,

$$h^{(3)}_{st} = \sum_{ijkl}\sum_{mnpq} d_{im}(\xi) d^{*}_{jn}(\xi) d_{kp}(\xi_1) d^{*}_{lq}(\xi_2) \times (i!j!k!l!m!n!p!q!)^{-1}\delta[s,i+j+k+l] \times \delta[t,m+n+p+q], \tag{4.24}$$

where each term has been defined previously. Despite the proliferation of indices, these terms have simple topological interpretations. The first describes a diagram where two chainlike terms join the unpaired dots, and do not end on one member of a pair. The second describes a diagram just like the first, but where one link in one of the two chainlike terms has been removed. The third describes a diagram in which one link in each of the two chainlike terms has been removed. On summing over all contributions consistent with the number of pairs of each type, the quoted results are obtained.

The remaining step is the collection of all terms in the variance expansion. From (4.3) and (2.14) we obtain

$$e^{K}\phi_1(\xi_1,z)\phi_1(\xi_2,z) = \sum_{qq'} R_{qq'} F_{qq'}, \tag{4.25}$$

where

$$F_{qq'} = \sum_{m,n} \frac{(\sigma_1\sigma_2 z)^{m+n+q+q'+2}}{(m+q+1)!(n+q'+1)!} \times (h^{(1)}_{mn} + h^{(2)}_{mn} + h^{(3)}_{mn}). \tag{4.26}$$

This leads to the following integral representation of the first term in δE^2. From (4.26), one obtains

$$e^{K}\phi_1\phi_1 = \int \frac{d\omega_1 d\omega_2}{(2\pi i)^2\omega_1\omega_2} e^{Q} F(\omega_1,\omega_2), \tag{4.27}$$

where Q is defined by (4.8) and

$$F = \sum_{qq'} F_{qq'} \omega_1^{-q}\omega_2^{-q'}. \tag{4.28}$$

The noise source term ψ_N has a similar diagrammatic expansion. The vacuum diagrams are identical to those just obtained but the connected pieces are different. The appearance of only two unmatched dots (the first pair) gives an expansion similar to the mean field case

$$h^{r}_{N\,st} = s!t![\delta[r,s+t+1]h^{(1)}_{N\,st} + \delta[r,s+t]\,h^{(2)}_{N\,st}]. \tag{4.29}$$

Here

$$h^{(1)\sigma}_{N\,st} = \int d\xi_\alpha \cdots d\xi_\beta G(\xi,\xi_\alpha)\cdots G(\xi_\beta,\xi) \tag{4.30}$$

and

$$h^{(2)}_{N\,st} = \sum_{ij} d_{s-i,t-j}(\xi) d^{*}_{ij}(\xi)(C^{s}_{i} C^{t}_{j}/S!t!), \tag{4.31}$$

where $d^{(3)}$ is a sum over permutations of the ordering of the integration limits. Again one obtains an integral representation for the noise source term. From (4.5)

$$\psi_N = \int \frac{d\omega_1 d\omega_2 e^{Q}}{(2\pi i)^2\omega_1\omega_2} F_N(\omega_{12}\omega_2), \tag{4.32}$$

where

$$F_N = \frac{N^2}{\sigma_1\sigma_2} G_N(\xi_1-\xi_2)\sum_{nm} \frac{(\sigma_1\sigma_2 z)^{n+m+1+q+q'}}{(n+q+1)!(m+q'+1)!} \times (h^{(1)}_{Nqq} + h^{(2)}_{Nqq'})\omega_1^{-q}\omega_2^{-q'}. \tag{4.33}$$

V. APPLICATION: INITIAL BEHAVIOR

The formalism developed in the preceding sections gives a general method of analyzing pump stochasticity. Under certain circumstances, and despite the profusion of mathematical symbols, the results for the Stokes wave may be expressed in a simple and useful form. The method is illustrated first by analyzing the initial behavior of the system. In the interests of simplicity the coherent part of the pump is taken to be a square pulse. Thus[20]

$$\chi(\xi,\xi_1) = I_c(\xi-\xi_1), \tag{5.1}$$

where I_c is the intensity of the coherent part of the pump alone. In computing the Stokes wave behavior, it is assumed that a small plane wave signal is incident at the point $z = 0$. These idealizations best illustrate the effect of pump noise.

The initial behavior of the system deals with the region $\xi - \xi' \lesssim \tau$, the correlation time of the noise component. In this region the correlation function is given by[20]

$$G(\xi,\xi') = I_r \tag{5.2}$$

and is independent of ξ. The various diagrams are trivial to compute, and if

$$\alpha = I_r(\xi-\xi_1), \tag{5.3}$$

the loops and chains of the field expansion are

$$L_n = \alpha^n, \tag{5.4a}$$

$$C_n = \kappa\alpha^n, \tag{5.4b}$$

where κ is the ratio of the coherent and noise intensities

$$\kappa = I_c/I_r. \tag{5.5}$$

From (3.27) the vacuum terms may be summed to give the exponent P,

$$P = \chi\omega + 2\kappa[(1-\alpha\omega)^{-1} - 1 - \alpha\omega] - \ln(1-\alpha\omega) \tag{5.6}$$

The series for P converges only if $|\alpha\omega| < 1$, thus restricting the contour of the complex integral, to avoid the singularities of P at $\alpha\omega = 1$. The connected diagrams are

$$h_s^{(1)} = I_r\alpha^s, \tag{5.7}$$

$$h_s^{(2)} = I_c(s+1)\alpha^s, \tag{5.8}$$

and from (3.29) the connected pieces may be found using the formula, valid for all x,y,

$$\sum_{nq} x^n y^q/(n+q+1)! = (x-y)^{-1}(e^x - e^y). \tag{5.9}$$

Thus,

$$f_1 = bI_r\left(1 + \kappa\frac{\partial}{\partial\alpha}\alpha\right)\left(b\alpha - \frac{b}{\omega}\right)^{-1}(e^{b\alpha} - e^{b/\omega}), \tag{5.10}$$

where $b = \sigma_1\sigma_2 z$. Inserting these expressions into (3.28), and noting the contour restrictions gives

$$e^K\phi_1(\xi_1,z) = bI_r\left(1 + \kappa\frac{\partial}{\partial\alpha}\alpha\right)J(b\alpha,\kappa), \tag{5.11}$$

where

$$J(x,y) = \int \frac{d\sigma}{2\pi i(1-\sigma)^2}\exp\left[\frac{x}{\sigma} + y\left(\frac{2}{1-\sigma} - 2 - \sigma\right)\right] \tag{5.12}$$

around a contour $|\sigma| < 1$. It may be written as a sum over Laguerre polynomials and Bessel functions,

$$J(x,y) = e^x\sum_{n=0}^{\infty}(x/2y)^{(n+1)/2}(-1)^n L_n^{(-1)}(-y)$$

$$I_{n+1}(2\sqrt{2xy}). \tag{5.13}$$

This result describes the initial behavior of the Stokes field until either $\xi \gtrsim \tau$ or the intensity is comparable to the pump intensity.

In the limit of no noise, $\alpha \to 0$, and $\kappa \to \infty$ but $\kappa b\alpha$ remains finite. Then it is straightforward to show that the purely coherent results (2.13) and (2.14) follow from (5.11) and (5.13). In the opposite extreme of pure noise, $\kappa \to 0$ and (5.13) gives a remarkably simple result,

$$e^K\phi_1 = bI_r\exp[bI_r(\xi - \xi')]. \tag{5.14}$$

From (2.13) the gain experienced by the Stokes field is exactly

$$\langle E(\xi,z)/E_0\rangle = \frac{G}{G-2}\{\exp[\tfrac{1}{2}\Gamma\xi(G-2)] - 1\} + 1. \tag{5.15}$$

Here G is the nominal gain

$$G = \gamma I_r z, \tag{5.16}$$

where γ is the steady-state gain coefficient for the Stokes intensity under coherent pumping conditions

$$\gamma = 2\sigma_1\sigma_2/\Gamma. \tag{5.17}$$

The field behaves initially in a manner reminiscent of a phase transition. For weak pumping, the field tends to a finite limit as ξ increases, but for strong pumping, the field grows exponentially.

The variance of the field may be found analogously. Defining

$$\alpha = I_r(\xi - \xi_1), \tag{5.18a}$$

$$\beta = I_r(\xi - \xi_2), \tag{5.18b}$$

gives the loops and chains of the variance expansion,

$$L_{kl} = \alpha^k\beta^l, \tag{5.19a}$$

$$C_{kl} = \kappa\alpha^k\beta^l \tag{5.19b}$$

From (4.8) the vacuum terms give the exponent

$$Q = \kappa f(\alpha\omega_1 + \beta\omega_2) - \ln(1 - \alpha\omega_1 - \beta\omega_2), \tag{5.20}$$

where

$$f(x) = 2(1-x)^{-1} - 2 - x. \tag{5.21}$$

The connected pieces are conveniently written

$$h_{st}^{(j)} = C^{(j)}\alpha^s\beta^t Y_{st}^{(j)}, \tag{5.22}$$

where

$$C^{(j)} = (2I_r^2, 3I_r I_s, I_c^2) \tag{5.23}$$

and

$$Y_{st}^{(j)} = C_s^{s+j}C_t^{j+t}. \tag{5.24}$$

Using (5.9) the function F, defined in (4.28), may be written in a closed form, from which the variance expansion follows,

$$e^K\phi_1(\xi_1,z)\phi_1(\xi_2,z) = \left[\frac{I_c^2}{36}\left(\frac{\partial^2}{\partial\alpha\partial\beta}\right)^3(\alpha\beta)^2 + \frac{3I_cI_r}{4} \times\left(\frac{\partial^2}{\partial\alpha\partial\beta}\right)^2\alpha\beta + 2I_r^2\frac{\partial^2}{\partial\alpha\partial\beta}\right]J, \tag{5.25}$$

where

$$J = \int \frac{d\sigma_1 d\sigma_2}{(2\pi i)^2(1-\sigma_1)(1-\sigma_2)(1-\sigma_1-\sigma_2)} \times\exp\left(\kappa f(\sigma_1+\sigma_2) + \frac{b\alpha}{\sigma_1} + \frac{b\beta}{\sigma_2}\right). \tag{5.26}$$

This integral does not have a useful repesentation in terms of the standard transcendental functions.

In the limit of no noise, (5.25) and (5.26) reduce to the square of the coherent pump result for the field. Consequently the variance vanishes,

$$\delta E^2 = 0. \tag{5.27}$$

In the case of pure noise, the integral J is easily evaluated to give

$$J = e^{b(\alpha+\beta)}[I_0(2b(\alpha\beta)^{1/2}) - 1]. \tag{5.28}$$

Inserting this into (5.25) gives the general result

$$\int d\xi_1 d\xi_2 \exp[-\Gamma(2\xi-\xi_1-\xi_2)]e^K\phi_1(\xi_1,z)\phi_1(\xi_2,z)$$

$$=\sum_{n=1}^{\infty}[C_n(G\Gamma\xi,\Gamma\xi)/n!]^2, \tag{5.29}$$

where

$$C_n(x,y)=e^{x-y}x^n[1+y\Phi(1,n+1;y-x)/n+1] \tag{5.30}$$

and Φ is a confluent hypergeometric function. The sum increases without limit as $\xi\to\infty$ if $G_r\geqslant 1$, otherwise it tends to a finite value. This qualitative behavior is similar to that of the field, but the threshold is smaller by a factor of 2. More detailed comparison with (5.14) shows that the variance is dominated by the first term in (4.3), and is much larger than the mean field squared. Thus the Stokes wave is mostly noise. That is, for $G\to\infty$

$$\delta E^2/E_0^2=2e^{2s}I_0(2s)-3e^{2s}+2e^s-1, \tag{5.31}$$

where $s=\frac{1}{2}G\Gamma\xi$, and for $G=2$

$$\delta E^2/E_0^2=4[I_0(2s)+I_1(2s)]-3s^2-4s-4, \tag{5.32}$$

while, for $G\ll 1$

$$\delta E^2/E_0^2=1/G^2. \tag{5.33}$$

Neglecting the contribution of the mean field squared to the mean intensity, the gain experienced by the Stokes intensity may be written approximately

$$\langle I_s/I_{in}\rangle\simeq\exp[2(2G-1)\Gamma\xi]+1. \tag{5.34}$$

Of the two extremes, pure noise and no noise, the former can be much more effective in producing gain at the Stokes frequency. The coherent result, obtained from (2.13), is as follows,

$$I_s/I_{in}=\exp[2G(1-\eta)], \tag{5.35}$$

where η is a function of $\Gamma\xi$ and G which tends to zero as $\xi\to\infty$ and which is negligible for $\Gamma\xi\gtrsim G$. The coherent case has a steady state limit, whereas the pure noise case increases with ξ until either $\xi>\tau$ or pump depletion is significant. For small ξ, the coherent case grows slowly in time,

$$I_s/I_{in}\simeq\exp[2(G\Gamma\xi)^{1/2}]. \tag{5.36}$$

Again, this is in marked contrast to the noise pump case.

VI. CHAIN APPROXIMATION

The correlation function for Gaussian noise is

$$G(\xi,\xi')=I_r\exp[-(\xi-\xi')^2/\tau^2], \tag{6.1}$$

where τ is the correlation time. The exact expressions for the loops and chains obtained using (6.1) are not simple in general, and do not allow an analytic solution. However, useful analytic expressions may be obtained for the experimentally important case where the noise component is small compared to the mean pump field. In this case the loops other than L_1 may be neglected compared to the chains; hence the name chain approximation.

The chains obtained from (6.1) are well approximated by the following:

$$C_n=\kappa\lambda(\alpha/\lambda)^n, \tag{6.2}$$

$$C_{kl}=\kappa\alpha^k\beta^l(1/2\mu^2+1/2\lambda^2)^{(k+l-1)/2}, \tag{6.3}$$

where α,β are given by (5.18) and λ,μ are defined by

$$\lambda^2=1+\pi^{-1}[(\xi-\xi')/\tau]^2, \tag{6.4}$$

$$\mu=\lambda(\xi,\xi'') \tag{6.5}$$

These expressions are least accurate for $\xi-\xi'\sim\tau$, where the error is small (a few percent), but away from that region the error is negligible. The mean pump intensity is $I_m=I_c+I_r$, for which the nominal gain for the intensity of the Stokes wave is

$$\bar{G}=\gamma I_m z. \tag{6.6}$$

The ripple gain is defined as in (5.16),

$$G_r=\gamma I_r z. \tag{6.7}$$

The Stokes field is given by (2.13) and (3.28). The complex integral is straightforwardly evaluated using the method of steepest descents. The field itself may then be found using a similar technique for the ξ-integral. In the transient regime the result in lowest order is

$$E_s\sim E_s^{(0)}\exp[(2\bar{G}\Gamma\xi)^{1/2}-\Gamma\xi+G_r\Gamma\xi/\lambda] \tag{6.8}$$

and the chain approximation is valid in the region

$$G_r\ll(\bar{G}/2\Gamma\xi)^{1/2}. \tag{6.9}$$

In the steady state regime, the Stokes field in lowest order is

$$E_s\sim E_s^{(0)}\exp[\bar{G}/2+\pi^{1/2}G_r\Gamma\tau/(1+g^2)^{1/2}], \tag{6.10}$$

where

$$g=2\pi^{1/2}\Gamma\tau/\bar{G}. \tag{6.11}$$

The chain approximation is valid in the region

$$G_r\ll(\bar{G}/2\Gamma\tau)^{1/2}. \tag{6.12}$$

Thus as the correlation time increases, the approximation fails and the initial system behavior results of the previous section apply.

The lowest order chain approximation is essentially the same one used by D'yakov, and Pasmanik and Friedman. Their results correspond to retaining no loops and only the four lowest order chain C_2, C_{02}, C_{11}, and C_{02}. The next higher order chain approximation may be obtained by expanding $P(\omega)$ in 3.27 in a power series in $\beta=I_r/I_m-I_r$, and truncating the chain sum at C_3. It is simple to include the loop term L_2, to obtain a slightly more accurate expression. The complex integral may then be evaluated using the method of stationary phase. The next higher order contribution to the gain exponent is found to be

$$-2(G_r/\bar{G})^2(2\bar{G}\Gamma\xi)^{3/2}\lambda^{-2} \tag{6.13}$$

in the transient regime, and

$$-\pi^{1/2}G_r^2 g/(1+g^2) \tag{6.14}$$

in the steady state regime, where g is given by (6.11)

The noise at the Stokes frequency is given by the variance and may be evaluated analogously to the field, using (2.13) and (4.27). In the transient regime the ratio of variance to intensity is given by the following,

$$\Delta = -\ln(1 - \delta E^2/\langle E^2\rangle)$$

$$= 2G_r\Gamma\xi/\lambda - 6(G_r/\bar{G}\lambda)^2(2\bar{G}\Gamma\xi)^{3/2} \quad (6.15)$$

and in the steady-state regime by

$$\Delta = \pi^{1/2}[G_r\Gamma\tau(1+g^2)^{-1/2} - 3G_r^2 g(1+g^2)^{-1}]. \quad (6.16)$$

It is immediately clear that unless $\Gamma\tau < 1/G_r$ the Stokes wave is mainly noise. Thus the condition for reproducibility of the experiment is

$$G_r\Gamma\tau \ll 1.$$

This condition does not seem to be a stringent requirement for molecular vibrational Raman scattering.

VII. LOOP APPROXIMATION

The chain approximation describes pumping by a sources with a strong monochromatic, i.e., nonstochastic component. In the other extreme of pure noise pumping, $A = 0$. Again, this case does not allow a simple analytic result in general, but good approximations are available. If $A = 0$, the chain terms vanish, resulting in the loop approximation.

The only surviving connected diagrams are $h_s^{(1)}$. In general, these diagrams are complicated analytic expressions, but are approximately given by

$$h_s^{(1)} \cong I_r(I_r\xi/\lambda)^s \exp[-\xi^2/\tau^2(s+1)]. \quad (7.1)$$

Thus even if $\xi \gg \tau$, if s is large enough, h_s may be significant. Since $P(\omega)$ is nonsingular at $\omega = 0$, only the singular part of the connected diagram sum is required. From (3.29), this is

$$f(\omega) = \omega e^{b/\omega}\sum_n \omega^n h_n^{(1)}$$

$$= \omega I_r e^{b/\omega}\sum_n (\tilde{\alpha}\omega)^n \exp[-\xi^2/\tau^2(s+1)], \quad (7.2)$$

where $\tilde{\alpha} = \alpha/\lambda$. In the limit $\omega\tilde{\alpha} \to 0$, the sum reduces to the lowest order term, but in the limit $\tilde{\alpha}\omega \to 1$ it gives a singular result,

$$\lim f_1 = \omega I_r e^{b/\omega}(1 - \tilde{\alpha}\omega)^{-1}. \quad (7.3)$$

If the complex integral is dominated by the region $\tilde{\alpha}\omega \sim 1$, then the higher order connected diagrams dominate the integral, otherwise only the lowest order diagram contributes. The loop diagrams may be evaluated in the region $\xi - \xi' \gg \tau$ to give the following contribution to (3.27)

$$P(\omega) = \alpha\omega\Phi\left(\tilde{\alpha}\omega, \tfrac{3}{2}, 1\right), \quad (7.4)$$

where Φ is a generalized Riemann function.[21]

Consider first broadband pumping $\Gamma\tau \ll 1$. Then the factor $b\tilde{\alpha} \ll \lambda$ and the complex integral is dominated by the region $\omega \to 0$. By use of the method of stationary phase, one obtains

$$e^K\phi_1 = (\lambda^3/4\pi)^{1/2}(b\tilde{\alpha})^{-1/4} b I_r G \exp(2G_r\Gamma\xi)^{1/2}. \quad (7.5)$$

A similar method applied to the ξ-integral gives the steady state gain,

$$E_s = E_s^{(0)} \exp\left[\tfrac{3}{4}(G_r\Gamma\tau)^{2/3}\right]. \quad (7.6)$$

This expression is valid for strong, broadband pumps

$$G_r\Gamma\tau \gg 1, \quad \Gamma\tau \ll (G_r/8)^{1/2}. \quad (7.7)$$

The variance expansion gives similar expressions, and the complex integrals are dominated by the region $\omega \to 0$. The loop sum is

$$Q = (\alpha\omega_1 + \beta\omega_2)\Phi(\alpha\omega_1/\lambda + \beta\omega_2/\mu, \tfrac{3}{2}, 1) \quad (7.8)$$

and the variance is extremely small,

$$\delta E^2/\langle E^2\rangle \approx 0(\Gamma\tau/G_r)^2. \quad (7.9)$$

Thus the intensity gain is twice the field gain, and the Stokes wave is not noise. This may indicate the possibility of converting broadband pumps to narrowband Stokes waves in the high gain region as well as the threshold region.[8]

For narrowband pumping $\Gamma\tau \gg 1$ and $b\tilde{\alpha} \gg \lambda$. Thus the complex integral is dominated by the region $\tilde{\alpha}\omega \sim 1$. Again, use of the method of stationary phase gives

$$e^K\phi_1 = (2\pi)^{-1/2}\exp\left[G_r\Gamma\xi/2\lambda + 1 + \zeta(\tfrac{3}{2})\lambda\right], \quad (7.10)$$

where ζ is the Riemann function. The steady state gain is

$$E_s = E_s^{(0)} \exp\sqrt{\pi}\,\Gamma\tau G_r(1 + \zeta(\tfrac{3}{2})/3\sqrt{\pi}\,\Gamma\tau) \quad (7.11)$$

which is valid in the region

$$\Gamma\tau \gg \zeta(\tfrac{3}{2})\pi^{-1/2}. \quad (7.12)$$

The variance expansion gives similar expressions and shows that the Stokes wave is noisy. The complex integral is dominated by the region $\tilde{\alpha}\omega_1 + \tilde{\beta}\omega_2 \sim 1$, and the variance is

$$\delta E^2/\langle E^2\rangle = 1 - \exp(-\sqrt{\pi}\,\Gamma\tau G_r). \quad (7.13)$$

The basic behavior of the system as the pump is broadened is clearly illustrated here. For an almost monochromatic noise pump where $\Gamma\tau \gg 1$, the Stokes wave grows exponentially in time according to the results of Sec. V. If pump depletion does not occur, a steady state configuration is reached where the gain is given by (7.11) and (7.13). This configuration is reached in a time $\sim G_r^{2/3}\tau$. As the pump is broadened the gain is reduced and the variance of the Stokes wave decreases. As $\Gamma\tau \ll 1$, the gain becomes nonlinear in the pump intensity and the Stokes wave narrows dramatically. This situation is given by (7.6). Qualitatively, these results agree with the predictions of D'yakov and Akhmanov. However, the boundary between broadband and narrowband pumps was not precisely known. It is given by (7.12).

VIII. DISCUSSION

The preceding sections have presented a method of analyzing pump stochasticity using functional analysis to generate a diagrammatic expansion for the expectation values of the Stokes wave amplitude and intensity. The meaning of this expectation value is defined in terms of an ensemble of experiments. It denotes the mean value of a quantity measured in a large number of experiments. The variance δE^2 is

the mean square deviation of the field amplitude in the ensemble; it is a direct measue of the reproducibility of the experiment. This gives a precise definition of noise as it is used here. It is also a critical parameter in evaluating both past experiments and the performance of future systems, as indicated in the Introduction.

The diagrammatic, systematic approach leads to representations of the mean quantities in terms of a contour integral and a real integral. From the point of view of obtaining general results, there are several ways to evaluate them; the most promising complex technique appears to be the method of steepest descents. In fact, the integrand is automatically in a form suitable for this. The real ξ-integral may be evaluated analogously, or by Gaussian quadrature, as suggested by the exponential damping factor. Although the results are not simple in general, under certain circumstances simple analytic expressions are found which can indicate much about the general behavior. A particularly interesting point arising out of the initial behavior results is the existence of an absolute instability, rather than the conventional convective instability found for the coherent pump case. For that region of the medium for which $G_r > 1$, the Stokes intensity grows exponentially in time. The instability is cut off either by pump depletion or by finite correlation time effects. For example, if the correlation time is very large, the expectation value of the Stokes intensity gorws until it is comparable to the mean pump intensity,

$$I_s \sim \exp 2G_r\Gamma\xi. \tag{8.1}$$

In this case the Stokes wave is mainly noise. Thus the absolute instability indicates that there is a large uncertainty as to the Stokes intensity in any particular experiment in the ensemble. The experiment is essentially irreproducible. On the other hand, if the correlation time is smaller so that

$$\tau \ll \bar{G}/2\Gamma G_r^2, \tag{8.2}$$

then the instability is no longer absolute, but convective. A quasi-steady-state solution exists in which the mean Stokes intensity approaches a finite limit as $\xi \to \infty$, but grows exponentially in z. The convective instability is cut off either by pump depletion or by the boundary of the medium. Note that the right-hand side of (8.2) varies as z^{-1} so that the longer the medium, the shorter the correlation time must be, to prevent the absolute instability. From the chain approximation results it is clear that the Stokes wave is not necessarily all noise. If the correlation time satisfies

$$\Gamma\tau \ll G_r, \tag{8.3}$$

then (6.13) shows that the Stokes wave has a small noise component and therefore the experiment is reproducible. Thus a small correlation time may bring the absolute instability to a convective one, and, in so doing, cause the experiment to become reproducible. Thus for $G_r \gtrsim 2$, the limits on the pump noise required for reproducibility are given by (8.2). Under these conditions the gain experienced by the Stokes intensity is

$$I_s \sim \exp[\bar{G} + 2G_r\pi^{1/2}\tau(1 + 4\pi(\Gamma\tau/\bar{G})^2)^{-1/2} + O(G_r^2\bar{G})] \tag{8.4}$$

in the steady state regime. This clearly demonstrates the role of the correlation time in mollifying and suppressing the potentially debilitating instabilities in the system.

Similar results apply to the case of pure noise pumping, where the Stokes wave is absolutely unstable for $\Gamma\tau \to \infty$, but as $\Gamma\tau$ decreases to the region

$$\Gamma\tau \ll (G_r/8)^{1/2} \tag{8.5}$$

the gain becomes nonlinear in the pump intensity, but the variance of the Stokes wave becomes very small. This may indicate that the Stokes wave is narrowband. The gain at the Stokes frequency is

$$I_s = I_s^{(0)} \exp\left[\tfrac{3}{2}(G_r\Gamma\tau)^{2/3}\right] \tag{8.6}$$

under these conditions.

The theory presented here is applicable to systems satisfying two important constraints. The first is the expression for the pump field (1.1), where the noise component B satisfies (1.2) and has no higher order correlation functions. This is, of course, the simplest way to introduce noise. However, higher order correlations undoubtedly exist, and, if desired, they may be incorporated into the diagrammatic approach as follows. In addition to the diagrams already discussed one must include more complicated ones where the dots are connected not by lines but are connected to several others, giving rise to star diagrams. This is best illustrated by an example. The fourth order correlation function appears in K defined in (2.11). It is represented by a diagram in which four lines emanate from a single point and end on dots, two from the upper row and two from the lower row. This four-point star is to be included with the line (i.e., a two-point star) in all possible combinations and permutations, leading to a series similar to the one for the two point star. Clearly the n,mth order correlation function is associated with an $(n + m)$-point star joining n dots from the upper row and m from the lower row. Although these expansions have been written down, a systematic study of them has not yet been carried out, in the belief that most of the important results are evident in the lowest order expansion.

The second constraint is in the restriction to plane waves. Not only are the spatial analogs of the temporal effects important, but there may be spatial effects with no temporal analogs; for example, transverse profile effects are germane to the study of amplified spontaneous emission. The value of the plane wave results lies in their rigor, and their potential for providing insight into more general systems. The spatial question will be the subject of further studies. Similar considerations apply to dispersive media, or the inclusion of a longitudinal spatial correlation length rather than a transverse one. It is likely that dispersion weakens the instability just as transverse stochasticity does. If the phase of the pump diffuses too rapidly for the molecular vibrations (or the idler wave) to follow it, then the instability is suppressed. One may speculate that this temporal phenomenon has spatial analogs. Defining a longitudinal coherence length L,

$$L = z\delta n, \tag{8.7}$$

where δn is the difference in refractive indices at the pump and Stokes frequency, and a transverse coherence time

$$\tau_2 = l_2/c\theta, \tag{8.8}$$

where l_2 is the transverse correlation length of the pump and θ is the root mean square angle between the pump and Stokes waves' directions of propagation, an effective correlation time is given by

$$1/\tau_{\text{eff}} = 1/\tau + 1/\tau_2 + c/L. \tag{8.9}$$

Replacing τ in (8.4) by τ_{eff} gives the gain which may be expected at the Stokes frequency. In the various limits discussed by Akhmanov[8] and by Pasmanik and Friedman,[9] this speculative prescription duplicates some of their results. It displays the suppression of the instabilities by longitudinal incoherence as discussed by Akhmanov,[8] and also the transverse effect observed by Zubarev *et al.*[14] Regardless of this speculation, the nondispersive plane wave case is expected to show the least controllable behavior, and forms an upper limit to the effects of pump stochasticity.

[1]Yu.E. D'yakov, Kratk. Soobshch. Fiz. **7**, 49 (1971).
[2]Yu.E. D'yakov, Kratk. Soobshch. Fiz. **12**, 41 (1971).
[3]Yu.E. D'yakov, Kratk. Soobshch. Fiz. **4**, 23 (1973).
[4]Yu.E. D'yakov, Kratk. Soobshch. Fiz. **5**, 39 (1973).
[5]Yu.E. D'yakov and A.I. Kovrigin, Sov. J. Quant. Electron. **5**, 1221 (1975).
[6]S.A. Akhmanov and Yu.E. D'yakov, Sov. Phys. JETP Lett. **18**, 305 (1973).
[7]S.A. Akhmanov and Yu.E. D'yakov, Sov. J. Quant. Electron. **2**, 362 (1973).
[8]S.A. Akhmanov, Yu.E. D'yakov, and L.I. Pavlov, Sov. Phys. JETP **39**, 249 (1974).
[9]G.A. Pasmanik and G.I. Friedman, Sov. J. Quant. Electron. **4**, 304 (1974).
[10]I.G. Zubarev and S.I. Mikhailov, Sov. J. Quant. Electron **4**, 351 (1974).
[11]I.G. Zubarev, A.B. Mironov, and S.I. Mikhailov, Sov. Phys. JETP Lett. **23**, 642 (1976).
[12]G.P. Dzhotyan, Yu.E. D'yakov, I.G. Zubarev. A.B. Mironov, and S.I. Mikhailov, Sov. J. Quant. Electron. **7**, 783 (1977).
[13]I.G. Zubarev and S.I. Mikhailov, Sov. J. Quant. Electron. **4**, 351, 683 (1974).
[14]I.G. Zubarev, A.B. Mironov, andn S.I. Mikhailov, Sov. J. Quant. Electron. **7**, 659 (1977).
[15]M.B. Vakhonev, V.N. Volkov, A.Z. Grasyuk, and A.N. Kirkin, Sov. J. Quant. Electron. **6**, 1369 (1976).
[16]A.Z. Grasyuk, Sov. J. Quant. Electron. **4**, 269 (1974).
[17]C.S. Wang, Phys. Rev. **182**, 482 (1969).
[18]V. Volterra, *Theory of Functionals and of Integral and Integrodifferential Equations* (Dove, New York, 1969).

[19]The work "vacuum" is borrowed from quantum field theory and indicates no connection to an external dot.
[20]We use units where $|A|^2 = I$.

[21]I.S. Gradshteyn and I.M. Rhyzik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), p. 1075.

Electrical conductivity of polycrystalline materials

M. W. M. Willemse and W. J. Caspers

Department of Applied Physics, Twente University of Technology, Enschede, The Netherlands
(Received 5 January 1979)

The problem of the determination of the macroscopic conductivity of statistically homogeneous and isotropic polycrystalline materials in terms of the principal values of the conductivity tensor of the constituent crystals is considered. A perturbation expansion, in terms of correlation functions with an optimal value of the zeroth-order estimate of the effective conductivity, is derived in which a separation is established into texture dependent and independent quantities. Consistence with the analogous expansion for the effective resistivity is demonstrated.

1. INTRODUCTION

In a polycrystal, a solid composed of randomly oriented crystalline regions, the orientation of the conductivity tensor varies with position whereas the principal values are everywhere the same. We restrict ourselves to variations of the orientation that are conveniently described from a statistical point of view: We exclude, for instance, periodic variations with position. Ensemble averages like the two-point correlation function $\langle \sigma_{ij}(\mathbf{r}_1)\sigma_{kl}(\mathbf{r}_2)\rangle$ are determined by choosing the same points $\mathbf{r}_1$ and $\mathbf{r}_2$ in each material sample and averaging over the ensemble of samples. Volume averaging is accomplished by taking the two points through the whole volume of the sample, which is eventually brought to infinity, keeping their relative position the same. With the help of an ergodic-type hypothesis we equate ensemble and volume averages, thus enforcing statistical homogeneity. The effective value of the conductivity tensor is defined by

$$\langle J_i\rangle = \sigma^*_{ij}\langle E_j\rangle, \tag{1.1}$$

or equivalently[1]

$$\langle J_i E_i\rangle = \sigma^*_{ij}\langle E_i\rangle\ \langle E_j\rangle. \tag{1.2}$$

Among the earlier proposals for the effective conductivity σ^* of a statistically isotropic polycrystal are the "parallel" effective value by Voigt,[2]

$$\sigma^* = \tfrac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3), \tag{1.3}$$

and the "series" value by Reuss,[3]

$$\sigma^* = \left[\frac{1}{3}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2} + \frac{1}{\sigma_3}\right)\right]^{-1}, \tag{1.4}$$

whereupon Hashin and Shtrikman[4] showed by using variational principles that these values actually serve as bounds. The use of a perturbation technique was proposed by Beran[1] and developed further by Molyneux,[5] Fokin,[6] and Kröner and Koch.[7] In Sec. 2 we give a modification of the procedure of Hori[8] giving a perturbation series as the formal solution, with this difference that he considers heterogeneous materials which are locally isotropic. Since it is known that σ^* lies between certain bounds, it is not obvious that we should start a perturbation expansion with one of these bounds as the zeroth-order estimate of σ^* but rather with a value in between, so for the time being we take an arbitrary value λ between these bounds as the zeroth-order value of σ^*: The notion "order of magnitude" permits this freedom of choice. For reasons of mathematical analogy the results are also valid for the effective dielectric constant, thermal conductivity, and magnetic permeability, and there is a parallel approach for mechanical properties of composite and fibre-reinforced materials. In Sec. 3 we calculate the low-order terms of the perturbation series, and in Sec. 4 we derive the corresponding expansion of the effective resistivity ρ^*. Partly new results for the bounds of σ^* are given in Sec. 5, and the final Sec. 6 concerns the derivation of a renormalized series for σ^*.

2. PERTURBATION SERIES

We start with Maxwell's equation in the case of stationary electrical conduction in the absence of an external magnetic field, with the continuity equation following from the conservation of charge, and with Ohm's law:

$$\mathrm{curl}\mathbf{E} = 0, \tag{2.1}$$

$$\mathrm{div}\mathbf{J} = 0, \tag{2.2}$$

$$\mathbf{J} = \boldsymbol{\sigma}\cdot\mathbf{E}. \tag{2.3}$$

The first equation permits the introduction of the electrical potential

$$\mathbf{E} = -\,\mathrm{grad}\Phi. \tag{2.4}$$

Combination of Eqs. (2.2), (2.3), and (2.4) gives a second-order partial differential equation with stochastic coefficients, which reads in indicial notation

$$\frac{\partial}{\partial x_i}\sigma_{ij}(\mathbf{r})\frac{\partial}{\partial x_j}\Phi(\mathbf{r}) = 0, \tag{2.5}$$

where Einstein's summation convention is employed, stating that summation over an index is implied when it appears twice.

Dividing the conductivity tensor into two parts

$$\sigma_{ij}(\mathbf{r}) = \lambda\,(\delta_{ij} + p_{ij}(\mathbf{r})), \tag{2.6}$$

such that λ does not depend on $\mathbf{r}$, we can write Eq. (2.5) as

 0022-2488/79/081824-08$01.00

$$\frac{\partial^2}{\partial x_i^2}\Phi(\mathbf{r}) = -\frac{\partial}{\partial x_i} p_{ij}(\mathbf{r}) \frac{\partial}{\partial x_j}\Phi(\mathbf{r}). \tag{2.7}$$

By using Green's function for the Laplacian in unlimited space we can formally solve this equation:

$$\Phi(\mathbf{r}_1) = \frac{1}{4\pi}\int_V d\mathbf{r}_2 \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \frac{\partial}{\partial x_{2,i}} p_{ij}(\mathbf{r}_2) \frac{\partial}{\partial x_{2,j}}\Phi(\mathbf{r}_2), \tag{2.8}$$

and after differentiating with respect to $\mathbf{r}_1$, we arrive at

$$E_i(\mathbf{r}_1) = \frac{1}{4\pi}\int_V d\mathbf{r}_2 \frac{x_{12,i}}{r_{12}^3} \frac{\partial}{\partial x_{2,j}} p_{jk}(\mathbf{r}_2) E_k(\mathbf{r}_2), \tag{2.9}$$

where we define: $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$.

The usual boundary condition $a(\mathbf{s})\Phi(\mathbf{s}) + b(\mathbf{s})\nabla_{\mathbf{n}}\Phi(\mathbf{s}) = 0$, with $\mathbf{s}$ on the bounding surface, is replaced by the condition that the average electric field $\langle \mathbf{E} \rangle$ is a given constant.

We assume the perturbation parameter $p_{ij} = (\sigma_{ij} - \lambda\delta_{ij})/\lambda$ to be small as compared to unity:

$$\sup_{\mathbf{r}} |p_{ij}(\mathbf{r})| \ll 1.$$

Next we expand the electric field intensity $\mathbf{E}$ in a series

$$E_i(\mathbf{r}) = \sum_{n=0}^{\infty} E_i^{(n)}(\mathbf{r}) = \langle E_i \rangle + \sum_{n=1}^{\infty} E_i^{(n)}(\mathbf{r}), \tag{2.10}$$

where each term $E_i^{(n)}$ is of the order of magnitude of the nth power of the perturbation parameter times the average electric field:

$$\mathbf{E}^{(n)}(\mathbf{r}) = o(|\mathbf{p}|^n)\langle \mathbf{E} \rangle. \tag{2.11}$$

Inserting Eq. (2.10) into Eq. (2.9) and equating terms of like order of magnitude, we get

$$E_i^{(n)}(\mathbf{r}_1) = \frac{1}{4\pi}\int_V d\mathbf{r}_2 \frac{x_{12,i}}{r_{12}^3} \times \frac{\partial}{\partial x_{2,j}} p_{jk}(\mathbf{r}_2) E_k^{(n-1)}(\mathbf{r}_2), \quad n = 1,2,\cdots. \tag{2.12}$$

Starting with $n = 1$,

$$E_i^{(1)}(\mathbf{r}_1) = \frac{1}{4\pi}\int_V d\mathbf{r}_2 \frac{x_{12,i}}{r_{12}^3} \frac{\partial}{\partial x_{2,j}} p_{jk}(\mathbf{r}_2)\langle E_k \rangle, \tag{2.13}$$

and changing the set of independent variables $\mathbf{r}_1$ and $\mathbf{r}_2$ into $\mathbf{r}_1$ and $\mathbf{r}_{12}$, we can write

$$E_i^{(1)}(\mathbf{r}_1) = \frac{1}{4\pi}\int_V d\mathbf{r}_{12} \frac{x_{12,i}}{r_{12}^3} \frac{\partial}{\partial x_{12,j}} p_{jk}(\mathbf{r}_1 + \mathbf{r}_{12})\langle E_k \rangle. \tag{2.14}$$

Similarly for $n = 2$,

$$E_i^{(2)}(\mathbf{r}_1) = \frac{1}{(4\pi)^2}\int_V d\mathbf{r}_2 \frac{x_{12,i}}{r_{12}^3} \frac{\partial}{\partial x_{2,j}} p_{jk}(\mathbf{r}_2) \times \int_V d\mathbf{r}_{23} \frac{x_{23,k}}{r_{23}^3} \frac{\partial}{\partial x_{23,l}} p_{lm}(\mathbf{r}_2 + \mathbf{r}_{23})\langle E_m \rangle, \tag{2.15}$$

and changing the set of independent variables $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_{23}$ into $\mathbf{r}_1, \mathbf{r}_{12}, \mathbf{r}_{23}$, we get

$$E_i^{(2)}(\mathbf{r}_1) = \frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23} \frac{x_{12,i}}{r_{12}^3} \frac{\partial}{\partial x_{12,j}} \frac{x_{23,k}}{r_{23}^3} \frac{\partial}{\partial x_{23,l}} \times p_{jk}(\mathbf{r}_1 + \mathbf{r}_{12}) p_{lm}(\mathbf{r}_1 + \mathbf{r}_{12} + \mathbf{r}_{23})\langle E_m \rangle, \tag{2.16}$$

and in general

$$E_i^{(n)}(\mathbf{r}_1) = \frac{1}{(4\pi)^n}\int d\mathbf{r}_{12}\cdots\int d\mathbf{r}_{n\,n+1} \times \frac{x_{12,i}}{r_{12}^3} \frac{\partial}{\partial x_{12,j}} \cdots \frac{x_{nn+1,m}}{r_{n\,n+1}^3} \frac{\partial}{\partial x_{n\,n+1,q}} \times p_{jk}(\mathbf{r}_1 + \mathbf{r}_{12})\cdots p_{qs}(\mathbf{r}_1 + \mathbf{r}_{12} + \cdots + \mathbf{r}_{n\,n+1})\langle E_s \rangle. \tag{2.17}$$

Assuming averaging to commute with differentiating and integrating, the average value of all terms of the electric field $\langle E_i^{(n)} \rangle$, $n \geqslant 1$, vanishes because the n-point correlation function $\langle p_{jk}(\mathbf{r}_2)\cdots p_{qs}(\mathbf{r}_{n+1}) \rangle$ is independent of $\mathbf{r}_{12}$, which makes Eq. (2.10) consistent.

The effective conductivity defined by Eq. (1.1) can now, by making use of Eqs. (2.6) and (2.10), be written as

$$\sigma_{ij}^* = \lambda\left(\delta_{ij} + \sum_{n=1}^{\infty} A_{ij}^{(n)}\right), \tag{2.18}$$

with

$$A_{ij}^{(1)} = \langle p_{ij}(\mathbf{r}) \rangle, \tag{2.19}$$

and for $n \geqslant 2$:

$$A_{ij}^{(n)} = \frac{1}{(4\pi)^{n-1}}\int d\mathbf{r}_{12}\cdots\int d\mathbf{r}_{n-1\,n} \times \frac{x_{12,k}}{r_{12}^3} \frac{\partial}{\partial x_{12,l}} \cdots \frac{x_{n-1\,n,s}}{r_{n-1\,n}^3} \frac{\partial}{\partial x_{n-1\,n,t}} \times \langle p_{ik}(\mathbf{r}_1) p_{lm}(\mathbf{r}_2)\cdots p_{qs}(\mathbf{r}_{n-1}) p_{tj}(\mathbf{r}_n) \rangle. \tag{2.20}$$

This series expansion (2.18) is similar to the one developed by Hori[8] with the difference that his series concerns a locally isotropic material.

For statistically isotropic materials all $A_{ij}^{(n)}$ reduce to a constant times a Kronecker delta so that we can write Eq. (2.18) in scalar form

$$\sigma^* = \lambda\left(1 + \sum_{n=1}^{\infty} A^{(n)}\right). \tag{2.21}$$

3. LOW-ORDER TERMS $A_{ij}^{(n)}$

The first-order perturbation term is simply

$$A_{ij}^{(1)} = \langle p_{ij} \rangle = \langle (\sigma_{ij} - \lambda\delta_{ij})/\lambda \rangle = [(\langle \sigma \rangle - \lambda)/\lambda]\delta_{ij}, \tag{3.1}$$

which vanishes when we choose $\lambda = \langle \sigma \rangle$.

In this case the second-order perturbation term as given in Eq. (2.20) reads

$$A_{ij}^{(2)} = \frac{1}{(4\pi)\langle \sigma \rangle^2}\int d\mathbf{r}_{12} \frac{x_{12,k}}{r_{12}^3} \frac{\partial}{\partial x_{12,l}} \langle \sigma'_{ik}(\mathbf{r}_1)\sigma'_{lj}(\mathbf{r}_1 + \mathbf{r}_{12}) \rangle. \tag{3.2}$$

This dimensionless second-rank tensor is a function of the two-point correlation function of the material. For the statistical description of the polycrystalline material we follow the cell model proposed by Miller[9]:

(1) The space is completely filled by nonoverlapping cells within which the material properties are constant,

(2) cells are distributed in a manner such that the material is statistically homogeneous,

(3) the material properties of a cell are statistically independent of those of any other cell,

(4) the material properties of a cell are statistically independent of the geometrical distribution (shape, size, orientation and arrangement) of cells.

Because of the independence of the orientation of the principal axes of the conductivity tensor on the geometry of the cells the correlation function appearing in Eq. (3.2) can be written as a product of a tensorial part and a spatial part:

$$\langle \sigma'_{ik}(\mathbf{r}_1)\sigma'_{ij}(\mathbf{r}_1+\mathbf{r}_{12})\rangle = \langle \sigma'_{ik}\,\sigma'_{ij}\rangle P(\mathbf{r}_{12}), \tag{3.3}$$

where $P(\mathbf{r}_{12})$ denotes the probability of finding the points $\mathbf{r}_1$ and $\mathbf{r}_2$ in the same cell.

For a statistically isotropic material we have $A^{(n)}_{ij} = A^{(n)}\delta_{ij}$ with $A^{(n)} = \frac{1}{3}A^{(n)}_{ii}$ so that Eq. (3.2) becomes

$$A^{(2)}_{ij} = \frac{1}{3}\delta_{ij}\frac{\langle \sigma'^2\rangle}{(4\pi)\langle\sigma\rangle^2}\int d\mathbf{r}_{12}\frac{x_{12,k}}{r_{12}^3}\frac{\partial}{\partial x_{12,k}}P(\mathbf{r}_{12}). \tag{3.4}$$

In spherical coordinates $x_k(\partial/\partial x_k)$ is simply $r(\partial/\partial r)$, whence

$$A^{(2)}_{ij} = \frac{1}{3}\delta_{ij}\frac{\langle \sigma'^2\rangle}{4\pi\langle\sigma\rangle^2}\int_0^{2\pi}d\phi\int_0^{\pi}d\theta\int_0^{\infty}dr\,r^2\sin\theta \times\frac{1}{r^3}\left(r\frac{\partial}{\partial r}\right)P(\mathbf{r}). \tag{3.5}$$

Assuming that there are no cells with infinite dimensions, so that

$$\lim_{r\to\infty}[P(\mathbf{r})] = 0,$$

we arrive at (see Appendix A)

$$A^{(2)}_{ij} = -\frac{1}{3}\frac{\langle \sigma'^2\rangle}{\langle\sigma\rangle^2}\delta_{ij} = -\frac{1}{3}\delta_2\delta_{ij}. \tag{3.6}$$

The third-order perturbation term of a statistically isotropic material is given by

$$A^{(3)} = \frac{1}{3}\frac{1}{(4\pi)^2\langle\sigma\rangle^3}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23} \times\frac{x_{12,k}}{r_{12}^3}\frac{\partial}{\partial x_{12,l}}\frac{x_{23,m}}{r_{23}^3}\frac{\partial}{\partial x_{23,n}} \times\langle \sigma'_{ik}(\mathbf{r}_1)\sigma'_{lm}(\mathbf{r}_1+\mathbf{r}_{12})\sigma'_{ni}(\mathbf{r}_1+\mathbf{r}_{12}+\mathbf{r}_{23})\rangle, \tag{3.7}$$

where again we chose $\lambda = \langle\sigma\rangle$.

Now we write the operators appearing in Cartesian form in Eq. (3.7) in terms of spherical harmonics:

$$\int d\mathbf{r}\frac{x_i}{r^3}\frac{\partial}{\partial x_j}f(\mathbf{r}) = \frac{1}{3}\delta_{ij}\int d\mathbf{r}\frac{1}{r^2}\frac{\partial}{\partial r}f(\mathbf{r}) + c_{ij,\mu}\int d\mathbf{r}\frac{1}{r^3} \times Y_{2\mu}(\hat{r})\left(1+\frac{1}{3}r\frac{\partial}{\partial r}\right)f(\mathbf{r}). \tag{3.8}$$

Assuming that all correlation functions do not depend on the direction of the vector connecting two consecutive points whenever these points are infinitely far apart we can write:

$$\frac{1}{4\pi}\int d\mathbf{r}\frac{x_i}{r^3}\frac{\partial}{\partial x_j}f(\mathbf{r}) = \frac{1}{3}\delta_{ij}[f(\infty)-f(0)] + \frac{1}{4\pi}c_{ij,\mu}\int d\mathbf{r}\frac{1}{r^3}Y_{2\mu}(\hat{r})f(\mathbf{r}). \tag{3.9}$$

TABLE I. The nonzero elements $c_{ij,\mu}$ connecting the Cartesian and spherical tensors.

$(\sqrt{5/6\pi})\,c_{ij,\mu}$	$\mu=-2$	$\mu=-1$	$\mu=0$	$\mu=1$	$\mu=2$
xx	1		$-\sqrt{2/3}$		1
xy	i				$-i$
xz		1		-1	
yx	i				$-i$
yy	-1		$-\sqrt{2/3}$		-1
yz		i		i	
zx		1		-1	
zy		i		i	
zz			$2\sqrt{2/3}$		

The constants $c_{ij,\mu}$ connecting Cartesian with spherical tensors[10] are symmetric and traceless in the indices ij and are given in Table I. Putting Eq. (3.9) into effect in Eq. (3.7), we get

$$A^{(3)} = \frac{1}{9}\frac{\langle\sigma'^3\rangle}{\langle\sigma\rangle^3} + \frac{1}{3}\frac{1}{(4\pi)^2\langle\sigma\rangle^3}c_{kl,\mu}c_{mn,\eta} \times\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}\frac{1}{r_{12}^3r_{23}^3}Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23}) \times\langle \sigma'_{ik}(\mathbf{r}_1)\sigma'_{lm}(\mathbf{r}_1+\mathbf{r}_{12})\sigma'_{ni}(\mathbf{r}_1+\mathbf{r}_{12}+\mathbf{r}_{23})\rangle, \tag{3.10}$$

and evaluating the three-point correlation function for all possible values of the indices k, l, m and n (see Appendix B), we arrive at

$$A^{(3)} = (\tfrac{1}{9}+\tfrac{1}{10}\alpha_3)\delta_3, \tag{3.11}$$

where we define

$$\alpha_3 = \frac{1}{3}c_{kl,\mu}c_{lk,\eta}\frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23} \times\frac{1}{r_{12}^3r_{23}^3}Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23})P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3), \tag{3.12}$$

with $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ denoting the probability of finding the three points in the same cell.

By using the addition theorem for spherical harmonics[10] and noting that $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$, because of statistical isotropy, only depends on r_{12}, r_{23} and $\theta = \cos^{-1}(\hat{r}_{12}\cdot\hat{r}_{23})$, Eq. (3.12) can be given a geometrical meaning:

$$\alpha_3 = \int_0^{\infty}\frac{dr_{12}}{r_{12}}\int_0^{\infty}\frac{dr_{23}}{r_{23}}\int_0^{\pi}d\theta\sin\theta\,\tfrac{1}{2}(3\cos^2\theta-1) \times P(r_{12},r_{23},\theta), \tag{3.13}$$

measuring the deviation of the cell form from a sphere.[7]

The fourth-order perturbation term as given in Eq. (2.20) reads

$$A^{(4)}_{ij} = \frac{1}{(4\pi)^3\langle\sigma\rangle^4}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}\int d\mathbf{r}_{34} \times\frac{x_{12,k}}{r_{12}^3}\frac{\partial}{\partial x_{12,l}}\frac{x_{23,m}}{r_{23}^3}\frac{\partial}{\partial x_{23,n}}\frac{x_{34,o}}{r_{34}^3}\frac{\partial}{\partial x_{34,p}} \times\langle \sigma'_{ik}(\mathbf{r}_1)\sigma'_{lm}(\mathbf{r}_2)\sigma'_{no}(\mathbf{r}_3)\sigma'_{pj}(\mathbf{r}_4)\rangle, \tag{3.14}$$

with $\lambda = \langle\sigma\rangle$.

By applying Eq. (3.9) to the variables $\mathbf{r}_{12}$, $\mathbf{r}_{23}$, and $\mathbf{r}_{34}$

and evaluating the four-point correlation function for all possible values of the indices and all configurations of the four points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, and $\mathbf{r}_4$ giving a nonvanishing contribution, Eq. (3.14) results for a statistically isotropic material in (see Appendices B and C):

$$A^{(4)} = \delta_4[-\tfrac{1}{27} - \tfrac{3}{10}\alpha_3 + \tfrac{1}{7}\alpha_{4,0}] + \delta_2^2[-\tfrac{1}{27} - \tfrac{4}{10}\alpha_3 + \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})], \tag{3.15}$$

which, because we are dealing with a one-phase polycrystalline material, can be written as (see Appendix A)

$$A^{(4)} = \delta_2^2[-\tfrac{5}{54} - \tfrac{1}{20}\alpha_3 + \tfrac{3}{14}\alpha_{4,0} + \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})], \tag{3.16}$$

where we define

$$\alpha_{4,0} = \tfrac{1}{3}c_{kl,\mu}c_{lm,\eta}c_{mk,\nu}(4\pi)^{-3}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23} \times \int d\mathbf{r}_{34}\,(r_{12}^3 r_{23}^3 r_{34}^3)^{-1} Y_{2\mu}(\hat{r}_{12}) Y_{2\eta}(\hat{r}_{23}) \times Y_{2\nu}(\hat{r}_{34}) P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4), \tag{3.17}$$

with $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$ the probability of finding the four points in the same cell, and

$$\alpha_{4,1} = \tfrac{1}{3}c_{kl,\mu}c_{lm,\eta}c_{mk,\nu}(4\pi)^{-3}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23} \times \int d\mathbf{r}_{34}\,(r_{12}^3 r_{23}^3 r_{34}^3)^{-1} Y_{2\mu}(\hat{r}_{12}) Y_{2\eta}(\hat{r}_{23}) Y_{2\nu}(\hat{r}_{34}) \times P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4), \tag{3.18}$$

with $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ the probability of finding the points $\mathbf{r}_1$ and $\mathbf{r}_2$ in one cell and the points $\mathbf{r}_3$ and $\mathbf{r}_4$ in another cell.

The factors $\alpha_{4,2}$ and $\alpha_{4,3}$ are defined by replacing $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ in the definition of $\alpha_{4,1}$ by $P(\mathbf{r}_1,\mathbf{r}_3;\mathbf{r}_2,\mathbf{r}_4)$ respectively $P(\mathbf{r}_1,\mathbf{r}_4;\mathbf{r}_2,\mathbf{r}_3)$.

The probability $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ introduces the notion of the exclusion principle: Since the probability of finding the four points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, and $\mathbf{r}_4$ in the same cell is already taken into account in $P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$, this probability must be excluded from $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$. When we denote the probability of finding the pair $(\mathbf{r}_1,\mathbf{r}_2)$ and the pair $(\mathbf{r}_3,\mathbf{r}_4)$ in one, possibly the same, cell by $P(\mathbf{r}_1,\mathbf{r}_2 \mid \mathbf{r}_3,\mathbf{r}_4)$, we can write

$$P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4) = P(\mathbf{r}_1,\mathbf{r}_2 \mid \mathbf{r}_3,\mathbf{r}_4) - P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4). \tag{3.19}$$

Calling probabilities like $P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3,\mathbf{r}_4)$ restricted[11] and probabilities like $P(\mathbf{r}_1,\mathbf{r}_2 \mid \mathbf{r}_3,\mathbf{r}_4)$ unrestricted, we can state that in all correlation functions we may replace restricted by unrestricted probabilities if at the same time we replace moments by cumulants.

4. DUAL PERTURBATION SERIES

Because the conductivity tensor and the resistivity tensor are locally related by $\sigma_{ik}\rho_{kj} = \delta_{ij}$, it is purely a matter of consistency to have the same relationship between their effective values:

$$\sigma_{ik}^*\rho_{kj}^* = \delta_{ij}. \tag{4.1}$$

So the determination of ρ_{ij}^* will lead to an equivalent mathematical problem.

Starting again with Eqs. (2.1), (2.2), and (2.3), we can write because of Eq. (2.2)

$$\mathbf{J} = \mathrm{curl}\psi. \tag{4.2}$$

Following the same procedure as in Sec. 2, using the free-space Green's function for the Laplacian, and dividing the resistivity tensor into two parts,

$$\rho_{ij}(\mathbf{r}) = \mu[\delta_{ij} + q_{ij}(\mathbf{r})], \tag{4.3}$$

we arrive at the following recursion relation:

$$J_i^{(n+1)}(\mathbf{r}_1) = \frac{-1}{4\pi}\int_V d\mathbf{r}_2 \epsilon_{ijk}\frac{x_{12,j}}{r_{12}^3}\epsilon_{klm}\frac{\partial}{\partial x_{2,l}} \times q_{mo}(\mathbf{r}_2) J_o^{(n)}(\mathbf{r}_2), \tag{4.4}$$

where ϵ_{ijk} is alternating Levi-Cevità tensor, having the property

$$\epsilon_{ijk}\epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}. \tag{4.5}$$

The effective resistivity is defined by

$$\langle E_i\rangle = \rho_{ij}^*\langle J_j\rangle. \tag{4.6}$$

Using Eq. (4.4), we can express ρ_{ij}^* in a perturbation series in terms of the current density:

$$\rho_{ij}^*\langle J_j\rangle = \mu\Big[\langle J_i\rangle + \langle q_{ij}\rangle\langle J_j\rangle + \sum_{n=1}^{\infty}\langle q_{ij}J_j^{(n)}\rangle\Big]. \tag{4.7}$$

So we arrive at a result analogous to the one given in Eqs. (2.18)–(2.20)

$$\rho_{ij}^* = \mu\Big(\delta_{ij} + \sum_{n=1}^{\infty} B_{ij}^{(n)}\Big), \tag{4.8}$$

with

$$B_{ij}^{(1)} = \langle q_{ij}\rangle, \tag{4.9}$$

$$B_{ij}^{(2)} = \frac{-1}{4\pi}\int d\mathbf{r}_{12}(\delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm}) \times \frac{x_{12,l}}{r_{12}^3}\frac{\partial}{\partial x_{12,m}}\langle q_{ik}(\mathbf{r}_1)q_{nj}(\mathbf{r}_2)\rangle, \tag{4.10}$$

$$B_{ij}^{(3)} = \Big(\frac{-1}{4\pi}\Big)^2\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}(\delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm}) \times (\delta_{os}\delta_{pt} - \delta_{ot}\delta_{ps})\frac{x_{12,l}}{r_{12}^3}\frac{\partial}{\partial x_{12,m}}\frac{x_{23,p}}{r_{23}^3}\frac{\partial}{\partial x_{23,s}} \times \langle q_{ik}(\mathbf{r}_1)q_{no}(\mathbf{r}_2)q_{tj}(\mathbf{r}_3)\rangle, \tag{4.11}$$

and for $B_{ij}^{(n)}$, $n \geqslant 4$, there exist expressions corresponding to $A_{ij}^{(n)}$ in the same way as $B_{ij}^{(2)}$ and $B_{ij}^{(3)}$ correspond to $A_{ij}^{(2)}$ and $A_{ij}^{(3)}$. Writing the Cartesian operator $x_i\,\partial/\partial x_j$ in terms of spherical harmonics as we did in Eq. (3.9), we get

$$\frac{-1}{4\pi}\int d\mathbf{r}(\delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm})\frac{x_l}{r^3}\frac{\partial}{\partial x_m}f(\mathbf{r}) = \frac{2}{3}\delta_{kn}[f(\infty) - f(0)] - \frac{1}{4\pi}c_{kn,\mu}\int d\mathbf{r}\times\frac{1}{r^3}Y_{2\mu}(\hat{r})f(\mathbf{r}). \tag{4.12}$$

This result exhibits the close relationship existing between the $A_{ij}^{(n)}$ and $B_{ij}^{(n)}$, which enables us to write down the $B_{ij}^{(n)}$ terms immediately, once we know the corresponding $A_{ij}^{(n)}$ terms. For a statistically isotropic polycrystal we have

$$B^{(1)} = \frac{1}{3}\langle q_{ii}\rangle = \frac{1}{3}\left\langle \frac{\rho_{ii} - \mu\delta_{ii}}{\mu}\right\rangle = 0, \tag{4.13}$$

$$B^{(2)} = -\frac{2}{3}\frac{\langle \rho'^2\rangle}{\langle\rho\rangle^2}, \tag{4.14}$$

$$B^{(3)} = \left(\tfrac{4}{9} + \tfrac{1}{10}\alpha_3\right)\frac{\langle\rho'^3\rangle}{\langle\rho\rangle^3}, \tag{4.15}$$

$$\begin{aligned} B^{(4)} &= \frac{\langle\rho'^4\rangle}{\langle\rho\rangle^4}\left(-\tfrac{8}{27} - \tfrac{6}{10}\alpha_3 - \tfrac{1}{7}\alpha_{4,0}\right) + \frac{\langle\rho'^2\rangle^2}{\langle\rho\rangle^4} \\ &\quad\times[\tfrac{4}{27} + \tfrac{8}{10}\alpha_3 - \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})] \\ &= \frac{\langle\rho'^2\rangle^2}{\langle\rho\rangle^4}[-\tfrac{8}{27} - \tfrac{1}{10}\alpha_3 - \tfrac{3}{14}\alpha_{4,0} \\ &\quad - \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})], \end{aligned} \tag{4.16}$$

where we have chosen $\mu = \langle\rho\rangle$.

5. INEQUALITIES AND BOUNDS

The perturbation terms $A_{ij}^{(n)}$ as introduced in Eq. (2.20) are defined by

$$\langle p_{ij}(\mathbf{r})E_j^{(n-1)}(\mathbf{r})\rangle = A_{ij}^{(n)}\langle E_j\rangle, \quad n\geqslant 1, \tag{5.1}$$

but can also be given by[12]

$$\langle E_i^{(n)}(\mathbf{r})E_i^{(m)}(\mathbf{r})\rangle = -A_{ij}^{(n+m)}\langle E_i\rangle\langle E_j\rangle, \quad n,m\geqslant 1, \tag{5.2}$$

and

$$\langle p_{ij}(\mathbf{r})E_i^{(n)}(\mathbf{r})E_j^{(m)}(\mathbf{r})\rangle = A_{ij}^{(n+m+1)}\langle E_i\rangle\langle E_j\rangle, \quad n,m\geqslant 0. \tag{5.3}$$

Similar expressions hold for the perturbation terms $B_{ij}^{(n)}$.

From these equations several inequalities concerning the perturbation terms $A_{ij}^{(n)}$ and $B_{ij}^{(n)}$ can be derived: Consider, for instance, the quadratic expression

$$\langle[E_i^{(n)}(\mathbf{r}) + E_i^{(n+1)}(\mathbf{r})]^2\rangle \geqslant 0, \tag{5.4}$$

from which we derive, using Eq. (5.2) in principal directions of $A_{ij}^{(n)}$,

$$A_i^{(2n)} + 2A_i^{(2n+1)} + A_i^{(2n+2)} \leqslant 0, \quad n\geqslant 1. \tag{5.5}$$

By using inequality (5.5) for $n = N, N+1, \cdots$ and summing, we deduce

$$A_i^{(2N)} + 2\sum_{n=2N+1}^{\infty} A_i^{(n)} \leqslant 0, \quad N\geqslant 1, \tag{5.6}$$

from which it follows

$$\sigma_i^* \leqslant \lambda\left(1 + \sum_{n=1}^{2N-1} A_i^{(n)} + \tfrac{1}{2}A_i^{(2N)}\right), \quad N\geqslant 1. \tag{5.7}$$

Doing the same thing with the dual series, we find bounds for the effective conductivity:

$$\begin{aligned} &\left(\mu\left[1 + \sum_{m=1}^{2M-1} B_i^{(m)} + \tfrac{1}{2}B_i^{(2M)}\right]\right)^{-1} \\ &\quad\leqslant \sigma_i^* \leqslant \lambda\left(1 + \sum_{n=1}^{2N-1} A_i^{(n)} + \tfrac{1}{2}A_i^{(2N)}\right), \quad M,N\geqslant 1. \end{aligned} \tag{5.8}$$

We do not intend to exploit this bounding expressing in this paper. Our aim is to find bounds for the form parameters introduced in Eqs. (3.12), (3.17), and (3.18) and, in order to derive the narrowest bounds possible, we will make an excursion to locally isotropic materials with random field $\{\sigma(\mathbf{r})\}$. Making use of the nonnegative character of the conductivity and of Eqs. (5.2) and (5.3), we can write:

$$\langle\sigma(\mathbf{r})E_i^{(1)}(\mathbf{r})E_i^{(1)}(\mathbf{r})\rangle \geqslant 0, \tag{5.9}$$

which for a statistically isotropic material leads to

$$-A^{(2)} + A^{(3)} \geqslant 0. \tag{5.10}$$

When we choose $\lambda = \langle\sigma\rangle$, we get, cf. Eqs. (3.6) and (3.11)

$$\tfrac{1}{3}\delta_2 + (\tfrac{1}{9} + \alpha_3)\delta_3 \geqslant 0, \tag{5.11}$$

which, together with the inequality, valid for a nonnegative stochastic variable,

$$\delta_2 + \delta_3 \geqslant 0, \tag{5.12}$$

leads for the extremum $\delta_3 = -\delta_2$ to an upper bound for α_3:

$$\alpha_3 \leqslant \tfrac{2}{9}. \tag{5.13}$$

In order to arrive at a lower bound, we consider the Schwarz inequality

$$\langle\sigma'^2(\mathbf{r})E_i^{(1)}(\mathbf{r})\rangle^2 \leqslant \langle\sigma'^2\rangle\langle\sigma'^2(\mathbf{r})E_i^{(1)}(\mathbf{r})E_i^{(1)}(\mathbf{r})\rangle, \tag{5.14}$$

which leads to (see Appendix C)

$$(-\tfrac{1}{3}\delta_3)^2 \leqslant \delta_2[\delta_4(\tfrac{1}{9} + \alpha_3) + \delta_2^2(\tfrac{2}{9} - \alpha_3)]. \tag{5.15}$$

In general it holds $\delta_3^2 \leqslant \delta_2\delta_4$ and $\delta_2^2 \leqslant \delta_4$. When we take $\delta_2^2 \ll \delta_4$ and $\delta_3^2 \simeq \delta_2\delta_4$, which is realizable for the random field $\{\sigma(\mathbf{r})\}$ of, for instance, a two-phase material, we find the lower bound for α_3 to be zero, so

$$0 \leqslant \alpha_3 \leqslant \tfrac{2}{9}. \tag{5.16}$$

Making use of the Schwarz inequality,

$$\langle E_i^{(1)}(\mathbf{r})E_i^{(2)}(\mathbf{r})\rangle^2 \leqslant \langle E_i^{(1)}(\mathbf{r})E_i^{(1)}(\mathbf{r})\rangle\langle E_i^{(2)}(\mathbf{r})E_i^{(2)}(\mathbf{r})\rangle, \tag{5.17}$$

we derive from Eq. (5.2) for a statistically isotropic material

$$(A^{(3)})^2 \leqslant A^{(2)}A^{(4)}. \tag{5.18}$$

For a locally isotropic material it follows, cf. Eqs. (3.11) and (3.15), that

$$\begin{aligned} [(\tfrac{1}{9} + \alpha_3)\delta_3]^2 \leqslant &-\tfrac{1}{3}\delta_2[\delta_4(-\tfrac{1}{27} - \alpha_3 + \alpha_{4,0}) \\ &+ \delta_2^2(-\tfrac{1}{27} + \alpha_3 + \alpha_{4,1} + \alpha_{4,2} + \alpha_{4,3})]. \end{aligned} \tag{5.19}$$

Taking the above-mentioned values for the normalized moments, we get an upper bound for $\alpha_{4,0}$

$$\alpha_{4,0} \leqslant \tfrac{1}{3}\alpha_3 - 3\alpha_3^2. \tag{5.20}$$

Following the same procedure for the dual perturbation terms $B^{(n)}$, we arrive at a lower bound for $\alpha_{4,0}$, which, together with the upper bound, leads to

$$-\tfrac{2}{3}\alpha_3 + \tfrac{3}{2}\alpha_3^2 \leqslant \alpha_{4,0} \leqslant \tfrac{1}{3}\alpha_3 - 3\alpha_3^2, \tag{5.21}$$

in agreement with Elsayed.[13]

Returning to our problem for locally anisotropic materials, we see that inequality (5.18) leads to an upper bound for the linear combination of form parameters $\alpha_{4,i}$, $i = 1,2,3$, as given in Eq. (3.16), whereas the corresponding inequality in terms of $B^{(n)}$ leads to a lower bound:

$$-\tfrac{4}{27} - \tfrac{1}{30}\alpha_3 + \tfrac{3}{400}\alpha_3^2 - \tfrac{3}{14}\alpha_{4,0}$$

$$\leqslant \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})$$
$$\leqslant \tfrac{2}{27} + \tfrac{1}{60}\alpha_3 - \tfrac{3}{200}\alpha_3^2 - \tfrac{3}{14}\alpha_{4,0}, \tag{5.22}$$

where we have taken the maximum value of $\delta_3^2/\delta_2^3 = \frac{1}{2}$.

Taking into account the relatively small contributions of α_3 and $\alpha_{4,0}$ because of Eqs. (5.16) and (5.21), we can approximate the bounds given in Eq. (5.22) by

$$-\tfrac{4}{27} \lesssim \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3}) \lesssim \tfrac{2}{27}. \tag{5.23}$$

6. RENORMALIZATION

We recall Eq. (3.9), which we wish to write symbolically as

$$x\frac{\partial}{\partial x} \to \frac{1}{3}R_\infty - \frac{1}{3}R_0 + Y_{2\mu}. \tag{6.1}$$

The operator R_∞ divides the correlation function into two parts if no cells with infinite dimensions are present in the ensemble, while the operator R_0 contracts two consecutive points in the correlation function.

When we evaluate the R_∞ operators in the $A_{ij}^{(n)}$ terms, we get new terms $D_{ij}^{(n)}$, in which only the operators R_0 and $Y_{2\mu}$ appear:

$$A_{ij}^{(n)} = A_{ij}^{(n)}(R_\infty, R_0, Y_{2\mu}), \tag{6.2}$$
$$D_{ij}^{(n)} = D_{ij}^{(n)}(R_0, Y_{2\mu}). \tag{6.3}$$

Expressions for the $A^{(n)}$ in terms of $D^{(n)}$ are

$$\begin{aligned} A_{ij}^{(1)} &= D_{ij}^{(1)}, \\ A_{ij}^{(2)} &= D_{ij}^{(2)} + \tfrac{1}{3}D_{ik}^{(1)}D_{kj}^{(1)}, \\ A_{ij}^{(3)} &= D_{ij}^{(3)} + \tfrac{1}{3}D_{ik}^{(1)}D_{kj}^{(2)} + \tfrac{1}{3}D_{ik}^{(2)}D_{kj}^{(1)} \\ &\quad + \tfrac{1}{9}D_{ik}^{(1)}D_{kl}^{(1)}D_{lj}^{(1)}, \\ &\quad \cdots, \end{aligned} \tag{6.4}$$

from which we derive

$$\begin{aligned} \sum_1^\infty A_{ij}^{(n)} &= \sum_1^\infty D_{ij}^{(n)} + \tfrac{1}{3}\sum_1^\infty D_{ik}^{(n)}\sum_1^\infty D_{kj}^{(n')} \\ &\quad + \tfrac{1}{9}\sum_1^\infty D_{ik}^{(n)}\sum_1^\infty D_{kl}^{(n')}\sum_1^\infty D_{lj}^{(n'')} + \cdots \\ &= \left(\frac{\Sigma_1^\infty D^{(n)}}{1 - \frac{1}{3}\Sigma_1^\infty D^{(n')}}\right)_{ij}. \end{aligned} \tag{6.5}$$

Evaluating the R_0 operator in $D_{ij}^{(n)}$, we get

$$\begin{aligned} D_{ij}^{(1)} &= \langle p_{ij}\rangle, \\ D_{ij}^{(2)} &= -\tfrac{1}{3}\langle p_{ij}^2\rangle + c_{kl,\mu}\frac{1}{4\pi} \\ &\quad \times \int d\mathbf{r}_{12}\frac{1}{r_{12}^3}Y_{2\mu}(\hat{r}_{12})\langle p_{ik}(\mathbf{r}_1)p_{lj}(\mathbf{r}_2)\rangle, \\ D_{ij}^{(3)} &= \tfrac{1}{9}\langle p_{ij}^3\rangle - \tfrac{1}{3}c_{kl,\mu}\frac{1}{4\pi}\int d\mathbf{r}_{12}\frac{1}{r_{12}^3}Y_{2\mu}(\hat{r}_{12})\langle p_{ik}(\mathbf{r}_1)p_{lj}^2(\mathbf{r}_2)\rangle \\ &\quad - \tfrac{1}{3}c_{kl,\mu}\frac{1}{4\pi}\int d\mathbf{r}_{12}\frac{1}{r_{12}^3}Y_{2\mu}(\hat{r}_{12})\langle p_{ik}^2(\mathbf{r}_1)p_{lj}(\mathbf{r}_2)\rangle \\ &\quad + \tfrac{1}{9}c_{kl,\mu}c_{mn,\eta}\frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23} \\ &\quad \times \frac{1}{r_{12}^3}\frac{1}{r_{23}^3}Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23}) \\ &\quad \times \langle p_{ik}(\mathbf{r}_1)p_{lm}(\mathbf{r}_2)p_{nj}(\mathbf{r}_3)\rangle, \end{aligned} \tag{6.6}$$

and so on.

Now we define the renormalized perturbation tensor:

$$\tilde{\sigma}_{ij} = p_{ij} - \tfrac{1}{3}p_{ij}^2 + \tfrac{1}{9}p_{ij}^3 \cdots = \left(\frac{p}{1 + \frac{1}{3}p}\right)_{ij}, \tag{6.7}$$

with principal values

$$\tilde{\sigma}_i = 3\frac{\sigma_i - \lambda}{\sigma_i + 2\lambda} \quad \text{(no summation)}. \tag{6.8}$$

Summing the sequence of Eqs. (6.6), we have

$$\sum_1^\infty D_{ij}^{(n)} = \sum_1^\infty T_{ij}^{(n')}, \tag{6.9}$$

with

$$\begin{aligned} T_{ij}^{(1)} &= \langle\tilde{\sigma}_{ij}\rangle, \\ T_{ij}^{(2)} &= c_{kl,\mu}\frac{1}{4\pi}\int d\mathbf{r}_{12}\frac{1}{r_{12}^3}Y_{2\mu}(\hat{r}_{12})\langle\tilde{\sigma}_{ik}(\mathbf{r}_1)\tilde{\sigma}_{lj}(\mathbf{r}_2)\rangle, \\ T_{ij}^{(3)} &= c_{kl,\mu}c_{mn,\eta}\frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}\frac{1}{r_{12}^3}\frac{1}{r_{23}^3}Y_{2\mu}(\hat{r}_{12}) \\ &\quad \times Y_{2\eta}(\hat{r}_{23})\langle\tilde{\sigma}_{ik}(\mathbf{r}_1)\tilde{\sigma}_{lm}(\mathbf{r}_2)\tilde{\sigma}_{nj}(\mathbf{r}_3)\rangle, \end{aligned} \tag{6.10}$$

and so on.

This result leads us to the renormalized series expansion of the effective conductivity:

$$\sigma_{ij}^* = \lambda\left[\frac{1 + \frac{2}{3}\Sigma_1^\infty T^{(n)}}{1 - \frac{1}{3}\Sigma_1^\infty T^{(n')}}\right]_{ij}, \tag{6.11}$$

with $T_{ij}^{(n)}$ defined in Eqs. (6.10).

As for the effective resistivity it is possible from Eq. (4.12) to derive an analogous formula. The renormalized perturbation tensor has the principal values:

$$\tilde{\rho}_i = \frac{q_i}{1 + \frac{2}{3}q_i} = 3\frac{\rho_i - \mu}{2\rho_i + \mu} = -3\frac{\sigma_i - 1/\mu}{\sigma_i + 2/\mu} = -\tilde{\sigma}_i, \quad \text{for } \mu = \frac{1}{\lambda}. \tag{6.12}$$

So the dual series can be written

$$\rho_{ij}^* = \frac{1}{\lambda}\left[\frac{1 - \frac{1}{3}\Sigma_1^\infty T^{(n)}}{1 + \frac{2}{3}\Sigma_1^\infty T^{(n')}}\right]_{ij}, \tag{6.13}$$

and it is concluded that the renormalized series given respectively in Eqs. (6.11) and (6.13) are consistent with Eq. (4.1).

The optimal value of λ may be determined by the condition that the sum of $T^{(n)}$ vanishes, but because the terms $T_{ij}^{(n)}$ depend on quantities concerning the random field $\{\sigma_{ij}(\mathbf{r})\}$ and on geometrical quantities concerning the texture, which are independent, it is not possible to make the sum of $T_{ij}^{(n)}$ equal to zero solely on the basis of knowledge of the random field $\{\sigma_{ij}(\mathbf{r})\}$. That is why the optimal value of λ follows from the condition that the trace of the first term $T_{ij}^{(1)}$ vanishes:

$$T_{ii}^{(1)} = \langle \tilde{\sigma}_{ii} \rangle = 3 \sum_i \frac{\sigma_i - \lambda}{\sigma_i + 2\lambda} = 0. \tag{6.14}$$

Equation (6.14) leads to a cubic equation for λ:

$$\lambda^3 - \tfrac{1}{4}(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_1\sigma_3)\lambda - \tfrac{1}{4}\sigma_1\sigma_2\sigma_3 = 0, \tag{6.15}$$

which can also be written in terms of fluctuations:

$$\lambda^3 - \tfrac{3}{4}\langle\sigma\rangle^2(1 - \tfrac{1}{2}\delta_2)\lambda - \tfrac{1}{4}\langle\sigma\rangle^3(1 - \tfrac{3}{2}\delta_2 + \delta_3) = 0, \tag{6.16}$$

with solution

$$\lambda = \langle\sigma\rangle(1 - \tfrac{1}{3}\delta_2 + \tfrac{1}{9}\delta_3 - \tfrac{5}{54}\delta_2^2\cdots). \tag{6.17}$$

Equation (6.14) leads also to a cubic equation for μ:

$$\mu^3 + (\rho_1 + \rho_2 + \rho_3)\mu^2 - 4\rho_1\rho_2\rho_3 = 0, \tag{6.18}$$

which is equivalent with Eq. (6.15) for $\lambda = 1/\mu$.

Equation (6.14) is equivalent with the solution of the effective medium theory[14,15] and Schulgasser's[16] Model 0.

The second term in Eq. (6.10), $T_{ij}^{(2)}$, vanishes because we are dealing with a statistically isotropic material, for which we can write $T^{(3)}$ and $T^{(4)}$ in terms of the form parameters defined in Eqs. (3.12), (3.17), and (3.18):

$$T^{(3)} = \tfrac{1}{10}\alpha_3\langle\tilde{\sigma}^3\rangle, \tag{6.19}$$

$$T^{(4)} = [\tfrac{3}{14}\alpha_{4,0} + \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})]\langle\tilde{\sigma}^2\rangle^2, \tag{6.20}$$

with $\langle\tilde{\sigma}^2\rangle = \delta_2 - \tfrac{2}{3}\delta_3 + \cdots$ and $\langle\tilde{\sigma}^3\rangle = \delta_3 - \tfrac{1}{2}\delta_2^2 + \cdots$.

Summarizing, we may state that the effective conductivity σ^* of a statistically isotropic polycrystalline material to fourth order is given by

$$\sigma^* = \lambda\{1 + \tfrac{1}{10}\alpha_3\langle\tilde{\sigma}^3\rangle + [\tfrac{3}{14}\alpha_{4,0} + \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3})]\langle\tilde{\sigma}^2\rangle^2\}, \tag{6.21}$$

with λ given by Eq. (6.15) and the form parameters α bounded by

$$\begin{aligned} &0 \leqslant \alpha_3 \leqslant \tfrac{2}{9}, \\ &-\tfrac{2}{3}\alpha_3 + \tfrac{3}{2}\alpha_3^2 \leqslant \alpha_{4,0} \leqslant \tfrac{1}{3}\alpha_3 - 3\alpha_3^2, \\ &-\tfrac{4}{27} \lesssim \tfrac{1}{100}(\alpha_{4,1} + 19\alpha_{4,2} + 10\alpha_{4,3}) \lesssim \tfrac{2}{27}. \end{aligned} \tag{6.22}$$

Let us finally apply the renormalization procedure to the two-dimensional case, where formally the same series expansion is valid, with the difference that factors $\frac{1}{3}$ are replaced by factors $\frac{1}{2}$ and the spherical harmonics $Y_{2\mu}$ are replaced by exponents $e^{i\mu\phi}$, $\mu = \pm 2$.

Because no rotationally invariant products of an odd number of traceless $\tilde{\sigma}_{ij}(\mathbf{r})$ components nor of an odd number of exponents $e^{i\mu\phi}$ can be made, all terms $T_{ij}^{(n)}$ $(n \geqslant 2)$ vanish while the condition $T_{ii}^{(1)} = 0$ immediately leads to the effective conductivity: $\sigma^* = (\sigma_1\sigma_2)^{1/2}$, in agreement with Dykhne[17] and Mendelson.[18]

APPENDIX A: CENTRAL MOMENTS

In order to promote a unified notation, we change the triplet of independent parameters $\{\sigma_1,\sigma_2,\sigma_3,\}$ into $\{\langle\sigma\rangle,\delta_2,\delta_3\}$, where δ_n is defined as the nth normalized moment:

$$\delta_n = \langle\sigma'^n\rangle/\langle\sigma\rangle^n, \quad n \geqslant 1. \tag{A1}$$

By means of the invariance equation

$$\sigma'^n_{ij} = \tfrac{3}{2}\langle\sigma'^2\rangle\sigma'^{n-2}_{ij} + \langle\sigma'^3\rangle\sigma'^{n-3}_{ij}, \tag{A2}$$

we see that all δ_n, $n \geqslant 4$, can be expressed in terms of δ_2 and δ_3, viz.

$$\delta_4 = \tfrac{3}{2}\delta_2^2, \quad \delta_5 = \tfrac{5}{2}\delta_2\delta_3, \quad \delta_6 = \tfrac{9}{4}\delta_2^3 + \delta_3^2, \quad \text{etc.} \tag{A3}$$

For the averages of the fluctuations of the resistivity tensor an analogous recursion relation exists. The connection between the two sets is expressed by

$$\langle\rho\rangle = \frac{1}{\langle\sigma\rangle}\left(\frac{1 - \frac{1}{2}\delta_2}{1 - \frac{3}{2}\delta_2 + \delta_3}\right), \tag{A4}$$

$$\frac{\langle\rho'^2\rangle}{\langle\rho\rangle^2} = \frac{\delta_2 - 2\delta_3 + \frac{1}{2}\delta_2^2}{(1 - \frac{1}{2}\delta_2)^2}, \tag{A5}$$

$$\frac{\langle\rho'^3\rangle}{\langle\rho\rangle^3} = -\frac{\delta_3 - \frac{3}{2}\delta_2^2 + \frac{3}{2}\delta_2\delta_3 + \frac{1}{4}\delta_2^3 - \delta_3^2}{(1 - \frac{1}{2}\delta_2)^3}. \tag{A6}$$

Extreme values for δ_2 and δ_3 are $\delta_2 = \delta_3 = 2$ for the limiting case of $\sigma_1 = \sigma_2 \ll \sigma_3$ and $\delta_2 = \frac{1}{2}$, $\delta_3 = -\frac{1}{4}$ for the limiting case of $\sigma_1 \ll \sigma_2 = \sigma_3$.

APPENDIX B: THE AVERAGE VALUE OF PRODUCTS OF TENSOR COMPONENTS

The average value of the product of two components of a second-rank tensor with random orientation is a component of an isotropic fourth-rank tensor:

$$\langle\sigma'_{ik}\sigma'_{lj}\rangle = a\delta_{ik}\delta_{lj} + b\delta_{il}\delta_{kj} + c\delta_{ij}\delta_{kl}. \tag{B1}$$

From the symmetry of the conductivity tensor it follows that $b = c$, and because the tensor is traceless we have

$$\langle\sigma'_{ii}\sigma'_{jj}\rangle = 9a + 6b = 0, \tag{B2}$$

so that

$$\langle\sigma'_{ik}\sigma'_{lj}\rangle = \tfrac{1}{10}\langle\sigma'^2\rangle[-2\delta_{ik}\delta_{lj} + 3(\delta_{il}\delta_{kj} + \delta_{ij}\delta_{kl})]. \tag{B3}$$

Proceeding in the same way for the product of three and four tensor components, we arrive at

$$\begin{aligned} &\langle\sigma'_{ik}\sigma'_{lm}\sigma'_{nj}\rangle \\ &= \tfrac{1}{70}\langle\sigma'^3\rangle\{16\delta_{ik}\delta_{lm}\delta_{nj} - 12[\delta_{ik}(\delta_{ln}\delta_{mj} + \delta_{lj}\delta_{mn}) \\ &\quad + \delta_{lm}(\delta_{in}\delta_{kj} + \delta_{ij}\delta_{kn}) + \delta_{nj}(\delta_{im}\delta_{kl} + \delta_{il}\delta_{km})] \\ &\quad + 9[\delta_{il}(\delta_{kn}\delta_{mj} + \delta_{kj}\delta_{mn}) + \delta_{im}(\delta_{kn}\delta_{lj} + \delta_{kj}\delta_{ln}) \\ &\quad + \delta_{in}(\delta_{kl}\delta_{mj} + \delta_{km}\delta_{lj}) + \delta_{ij}(\delta_{kl}\delta_{mn} + \delta_{km}\delta_{ln})]\}, \end{aligned} \tag{B4}$$

and

$$\begin{aligned} &\langle\sigma'_{ik}\sigma'_{lm}\sigma'_{no}\sigma'_{pj}\rangle \\ &= \tfrac{1}{70}\langle\sigma'^4\rangle\{4\delta_{ik}\delta_{lm}\delta_{no}\delta_{pj} - 2[\delta_{ik}\delta_{lm}(\delta_{np}\delta_{oj} \\ &\quad + \delta_{nj}\delta_{op}) + \delta_{ik}\delta_{no}(\delta_{lp}\delta_{mj} + \delta_{lj}\delta_{mp}) \\ &\quad + \delta_{ik}\delta_{pj}(\delta_{ln}\delta_{mo} + \delta_{lo}\delta_{mn}) \\ &\quad + \delta_{lm}\delta_{no}(\delta_{ip}\delta_{kj} + \delta_{ij}\delta_{kp}) \\ &\quad + \delta_{lm}\delta_{pj}(\delta_{in}\delta_{ko} + \delta_{io}\delta_{kn}) \end{aligned}$$

$$+\delta_{no}\delta_{pj}(\delta_{il}\delta_{km}+\delta_{im}\delta_{kl})]$$
$$+3[(\delta_{il}\delta_{km}+\delta_{im}\delta_{kl})(\delta_{np}\delta_{oj}+\delta_{nj}\delta_{op})$$
$$+(\delta_{in}\delta_{ko}+\delta_{io}\delta_{kn})(\delta_{lp}\delta_{mj}+\delta_{lj}\delta_{mp})$$
$$+(\delta_{ip}\delta_{kj}+\delta_{ij}\delta_{kp})(\delta_{ln}\delta_{mo}+\delta_{lo}\delta_{mn})]\}. \quad \text{(B5)}$$

APPENDIX C: EVALUATION OF THE FOURTH-ORDER PERTURBATION TERM $A_{ij}^{(4)}$

Elaborating the Cartesian operators $x\partial/\partial x$ in terms of spherical harmonics, as given by Eq. (3.9), in the fourth-order perturbation term $A_{ij}^{(4)}$ of a statistically isotropic material leads to

$$A_{ii}^{(4)}\langle\sigma\rangle^4 = -\frac{1}{27}\langle\sigma'^4_{ii}\rangle+\frac{1}{27}\langle\sigma'^2_{im}\rangle^2$$
$$-\frac{2}{3}c_{kl,\mu}c_{mn,\eta}\frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}\frac{1}{r_{12}^3r_{23}^3}$$
$$\times Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23})\langle\sigma'_{ik}(\mathbf{r}_1)\sigma'_{lm}(\mathbf{r}_2)\sigma'^2_{ni}(\mathbf{r}_3)\rangle$$
$$-\frac{1}{3}c_{kl,\mu}c_{op,\eta}\frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}$$
$$\times\int d\mathbf{r}_{23}\frac{1}{r_{12}^3r_{23}^3}Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23})$$
$$\times\langle\sigma'_{ik}(\mathbf{r}_1)Y\sigma'^2_{lo}(\mathbf{r}_2)\sigma'_{pi}(\mathbf{r}_3)\rangle$$
$$+c_{kl,\mu}c_{mn,\eta}c_{op,\nu}\frac{1}{(4\pi)^3}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}\int d\mathbf{r}_{34}$$
$$\times\frac{1}{r_{12}^3r_{23}^3r_{34}^3}Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23})Y_{2\nu}(\hat{r}_{34})$$
$$\times\langle\sigma'_{ik}(\mathbf{r}_1)\sigma'_{lm}(\mathbf{r}_2)\sigma'_{no}(\mathbf{r}_3)\sigma'_{pi}(\mathbf{r}_4)\rangle. \quad \text{(C1)}$$

In the third term we have

$$\langle\sigma'_{ik}(\mathbf{r}_1)\sigma'_{lm}(\mathbf{r}_2)\sigma'^2_{ni}(\mathbf{r}_3)\rangle$$
$$=\langle\sigma'_{ik}\sigma'_{lm}\sigma'^2_{ni}\rangle P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$$
$$+\langle\sigma'_{ik}\sigma'_{lm}\rangle\langle\sigma'^2_{ni}\rangle P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3), \quad \text{(C2)}$$

with

$$P(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_3)=P(\mathbf{r}_1,\mathbf{r}_2\mid\mathbf{r}_3)-P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$$
$$=P(\mathbf{r}_1,\mathbf{r}_2)-P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3), \quad \text{(C3)}$$

of which the term $P(\mathbf{r}_1,\mathbf{r}_2)$ vanishes upon integrating.

In the fourth term we have analogously

$$\langle\sigma'_{ik}(\mathbf{r}_1)\sigma'^2_{lo}(\mathbf{r}_2)\sigma'_{pi}(\mathbf{r}_3)\rangle$$
$$=(\langle\sigma'_{ik}\sigma'^2_{lo}\sigma'_{pi}\rangle-\langle\sigma'_{ik}\sigma'_{pi}\rangle\langle\sigma'^2_{lo}\rangle)$$
$$\times P(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)+\langle\sigma'_{ik}\sigma'_{pi}\rangle\langle\sigma'^2_{lo}\rangle P(\mathbf{r}_1,\mathbf{r}_3). \quad \text{(C4)}$$

With the help of Eq. (5.2) we deduce from

$$\langle E_i^{(1)}(\mathbf{r}_2)E_i^{(1)}(\mathbf{r}_2)\rangle=-A_{ij}^{(2)}\langle E_i\rangle\langle E_j\rangle \quad \text{(C5)}$$

that the following holds on the basis of Eq. (3.9):

$$\frac{1}{3}c_{ik,\mu}c_{im,\eta}\frac{1}{(4\pi)^2}\int d\mathbf{r}_{12}\int d\mathbf{r}_{23}\frac{1}{r_{12}^3r_{23}^3}$$
$$\times Y_{2\mu}(\hat{r}_{12})Y_{2\eta}(\hat{r}_{23})\langle\sigma'_{kl}(\mathbf{r}_1)\sigma'_{ml}(\mathbf{r}_3)\rangle=\tfrac{2}{9}\langle\sigma'^2\rangle. \quad \text{(C6)}$$

Now we are able to write $A^{(4)}$ in terms of the form parameters as defined in Sec. 3:

$$A^{(4)}=-\tfrac{1}{27}\delta_4+\tfrac{1}{27}\delta_2^2-\tfrac{2}{3}\tfrac{1}{10}\alpha_3(\delta_4-\delta_2^2)-\tfrac{1}{3}\alpha_3(\tfrac{7}{10}\delta_4-\delta_2^2)$$
$$-\tfrac{1}{3}\tfrac{2}{9}\delta_2^2+\tfrac{1}{7}\alpha_{4,0}\delta_4+\tfrac{1}{100}\delta_2^2(\alpha_{4,1}+19\alpha_{4,2}+10\alpha_{4,3}). \quad \text{(C7)}$$

[1] M.J. Beran, *Statistical Continuum Theories* (Interscience, New York, 1968).
[2] W. Voigt, *Lehrbuch der Kristallphysik* (Teuber, Leipzig, 1928), p. 410.
[3] A. Reuss and Z. Angew, Math. Mech. **9**, 49 (1929).
[4] Z. Hashin and S. Shtrikman, Phys. Rev. **130**, 129 (1963).
[5] J.E. Molyneux, J. Math. Phys. **11**, 1172 (1970).
[6] A.G. Fokin, Zh. Tekh. Fiz. **41**, 1073 (1971) [Sov. Phys. Tech. Phys. **16**, 849 (1971)].
[7] E. Kröner and H. Koch, SM **1**, 183 (1976).
[8] M. Hori, J. Math. Phys. **14**, 514 (1973).
[9] M.N. Miller, J. Math. Phys. **10**, 1988 (1969).
[10] A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1968).
[11] M. Hori and F. Yonezawa, J. Math. Phys. **15**, 2177 (1974).
[12] M. Hori, J. Math. Phys. **14**, 1942 (1973).
[13] M.A. Elsayed, J. Math. Phys. **15**, 2001 (1974).
[14] D.A.G. Bruggeman, Ann. Physik **25**, 645 (1936).
[15] M. Hori, J. Math. Phys. **18**, 487 (1977).
[16] K. Schulgasser, J. Appl. Phys. **47**, 1880 (1976).
[17] A.M. Dykhne, Zh. Eksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
[18] K.S. Mendelson, J. Appl. Phys. **46**, 917 (1975).

Erratum: The Feynman maps and the Wiener integral [J. Math. Phys. 19, 1742 (1978)]

A. Truman

Math Department, Heriot-Watt University, Edinburgh, Scotland

The proof of Theorem 6 in the above paper contains a mistake. I am grateful to David Elworthy of the Math Institute at Warwick University for pointing out the error. Substantially the same result is contained in Theorem 6′ below. This similar result is strong enough to validate the assertions following Theorem 6 in the original paper. We require the following lemma.

Lemma: Let $(1+K): H\to H$ be a bounded linear surjection. Let $P_m: H\to H$ be an orthogonal projection for $m = 1,2,\cdots$. Define the orthogonal surjective projection Q_m $: H\to(1+K)P_mH$, for $m = 1,2,\cdots$. Then, if $P_m \overset{s}{\to} 1$, $Q_m \overset{s}{\to} 1$, as $m\to\infty$, $1: H\to H$ being the identity.

Proof: Let $\gamma'\in H$ be given by $\gamma' = (1+K)\gamma$, for some $\gamma\in H$. Then $\gamma' = (1+K)P_m\gamma + (1+K)(1-P_m)\gamma$. Hence,

$$Q_m\gamma' = (1+K)P_m\gamma + Q_m(1+K)(1-P_m)\gamma. \tag{1}$$

$$\therefore \|Q_m\gamma' - \gamma'\| \leqslant 2\|1+K\|\,\|P_m\gamma - \gamma\| \to 0, \text{ as } m\to\infty, \quad \forall\ \gamma'\in H.$$

Theorem 6′: Let $(1+K): H\to H$ be a linear injection with K trace class and $\det(1+K)\neq 0$, det being the Fredholm determinant. Let $g: H\to\mathbb{C}$ and define $g_{1+K}: H\to\mathbb{C}$ by $g_{1+K}[\gamma] = g[(1+K)\gamma]$. It is convenient to denote by $(e_s^K g_{1+K})$

$$(e_s^K g_{1+K})[\gamma] = \exp\left(\frac{i}{2s}(K\gamma,K\gamma) + \frac{i}{s}(K\gamma,\gamma)\right) g_{1+K}[\gamma], \tag{2}$$

so that $(e_s^K g_{1+K}): H\to\mathbb{C}$. Then, if $g\in\mathcal{F}(H)$, $(e_s^K g_{1+K})\in\mathcal{F}^s(P_\infty H)$ and

$$\mathcal{F}^s[e_s^K g_{1+K}] = |\det(1+K)|^{-1}\mathcal{F}^s[g], \quad \mathrm{Im}s\leqslant 0, \quad s\neq 0. \tag{3}$$

Proof: Substitution for $(e_s^K g_{1+K})$ gives, for $m = 1,2,\cdots$,

$$\mathcal{F}^s_m[e_s^K g_{1+K}] = N_m\int \exp\left(\frac{i}{2s}\|Q_m\gamma'\|^2\right) g[Q_m\gamma']\, d(P_m\gamma), \tag{4}$$

where $Q_m: H\to(1+K)P_mH$ is the surjective orthogonal projection, $\gamma' = (1+K)P_m\gamma$ and $N_m = (2\pi ist/m)^{-m/2}$, the principal branch of the square root being taken.

We now change integration variables from $d(P_m\gamma)$ to $d(Q_m\gamma')$. Choose $e_i'\in P_mH$ such that $f_i = (1+K)e_i'$, $i = 0,1,2,\ldots,m-1$, form an orthonormal basis for $(1+K)P_mH$ Here, for $i = 0,1,\ldots,m-1$,

$$e_i' = \sum_{j=0}^{m-1} A_{ij}e_j,$$

with

$$e_j(\cdot) = \left[G\left(\frac{\overline{j+1}\,t}{m},\cdot\right) - G\left(\frac{jt}{m},\cdot\right)\right]\left(\frac{m}{t}\right)^{1/2},$$

$j = 0,1,2,\ldots,m-1$, $G(\sigma,\tau) = t - \sup\{\sigma,\tau\}$, the e_j's forming an orthonormal basis for P_mH. Then the A_{ij}'s must satisfy

$$\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} A_{ik}((1+K)e_k,(1+K)e_l)A_{jl} = \delta_{ij}, \tag{5}$$

$i,j = 0,1,2,\ldots,m-1$. The A_{ij}'s exist because the matrix with entries $((1+K)e_k,(1+K)e_l)$ is symmetric and positive definite.

Taking determinants we arrive at

$$|\det A| = |\det(1_m + P_mKP_m + P_mK^*P_m + P_mK^*KP_m)|^{-1/2}. \tag{6}$$

This is just the Jacobian determinant for the change of integration variables. To see this write

$$\gamma' = \left(\frac{m}{t}\right)^{1/2}\sum_{j=0}^{m-1}\tilde{\gamma}_j f_j, \quad P_m\gamma = \left(\frac{m}{t}\right)^{1/2}\sum_{k=0}^{m-1}\Delta\gamma_k e_k, \tag{7}$$

giving, for $j = 0,1,2,\ldots,m-1$,

$$\tilde{\gamma}_j = \sum_{k=0}^{m-1}(f_j,(1+K)e_k)\Delta\gamma_k = \sum_{k=0}^{m-1}(A^{-1})_k\Delta\gamma_k, \tag{8}$$

A^{-1} being the matrix inverse of A. Since $(1+K)e_0, (1+K)e_1, \ldots,(1+K)e_{m-1}$ are linearly independent vectors, $(1+K): H\to H$ being an injection, we obtain $0 < |\det A| < \infty$.

Thus, changing integration variables, for $\mathrm{Im}s < 0$, leads to

$$\mathcal{F}^s_m[e_s^K g_{1+K}] = N_m|\det(1+P_mLP_m)|^{-1/2} \times\int \exp\left(\frac{i}{2s}\|Q_m\gamma'\|^2\right) g[Q_m\gamma']\, d(Q_m\gamma'), \tag{9}$$

where L is the trace-class operator $L = (K + K^* + K^*K)$, det denotes the Fredholm determinant, and the integration variables $d(Q_m\gamma')$ denote $\Pi_{j=0}^{m-1}d\tilde{\gamma}_j$.

When $\mathrm{Im}s < 0$, for $g\in\mathcal{F}(H)$,

$$g[\gamma] = \int_H \exp[i(\gamma',\gamma)]\, d\mu_g(\gamma'),$$

arguing as in Theorem 3, we obtain

$$\mathcal{F}^s_m[e_s^K g_{1+K}] = |\det(1+P_mLP_m)|^{-1/2} \times\int \exp\left(\frac{-is}{2}(\gamma,Q_m\gamma)\right) d\mu_g(\gamma). \tag{10}$$

 0022-2488/79/081832-02$00.00

Letting $m \to \infty$, since det is trace-class continuous and $P_m L P_m \to L$ in trace norm as $m \to \infty$, the result for $\mathrm{Im} s < 0$ follows from the dominated convergence theorem for the measure μ_g and the factorization properties of det.

When $\mathrm{Im} s = 0$, $s \neq 0$, the justification of the change of integration variables and the validity of the last equation require closer attention. The required result for $\mathrm{Im} s = 0$ follows by looking more carefully at our definition of $\mathscr{F}^s_m$. Explicitly we have

$$\mathscr{F}^s_m[f] = N_m \lim_{R \to \infty} \int_{-R}^{R} \cdots \int_{-R}^{R} (f e_s \circ P_m)[\gamma]\, d(P_m \gamma), \quad (11)$$

where the finite limits of integration $-R$ and $+R$ refer to the integration variables $\Delta\gamma_j, j = 0,1,2,\ldots,m-1$ and the complex Gaussian $e_s[\gamma] = \exp[(i/2s)\|\gamma\|^2]$, $e_s : H \to \mathbb{C}$. The result for $ims = 0$ now follows in much the same way as in the proof of Theorem 3. (For further details see Ref. 1 where a more general result is proved.)

Theorem 7 requires the additional hypothesis: For a.e. real s, $f(s) = \lim_{\epsilon \downarrow 0} f(s - i\epsilon)$.

[1] K.D. Elworthy and A. Truman, "The Feynman maps and the anharmonic oscillator," in preparation.

Erratum: A nonlinear scalar field theory in isotropic homogeneous space–time [J. Math. Phys. 19, 2253 (1978)]

J. A. Okolowski
221 Irving Street, Chester, Pennsylvania 19013

Equations (1.4) and (3.10) are printed with typographical errors; the corrections are as follows: Eq. (1.4) should read

$$\frac{1}{R}\frac{\partial}{\partial t}\left(R^3\frac{\partial\Psi}{\partial t}\right)-\frac{1}{\sinh^2\chi}\frac{\partial}{\partial\chi}\left(\sinh^2\chi\frac{\partial\Psi}{\partial\chi}\right)$$
$$-\frac{1}{\sinh^2\chi\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Psi}{\partial\theta}\right)+\frac{1}{\sin\theta}\frac{\partial^2\Psi}{\partial\phi^2}\right]$$
$$+R^2\frac{\partial f}{\partial\Psi}=0,$$

and Eq. (3.10) should read

$$u(0)=0,\quad \lim_{\chi\to\infty}u(\chi)=0.$$

 0022-2488/79/081834-01$00.00