Topological solution for systems of simultaneous linear equations[a)]

Adel F. Antippa
Département de Physique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada G9A-5H7

Nguyen Ky Toan
Département de Mathématiques, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada G9A-5H7

(Received 13 June 1979; accepted for publication 21 August 1979)

Using the discrete path approach, we derive an explicit topological expression for the solution of a system of simultaneous linear equations. This is done by establishing a homomorphism between the solution x_n, and the set of paths, defined on the corresponding signal flow graph, from all sources to vertex n.

I. INTRODUCTION

The discrete path formalism[1] was first developed to solve finite-difference equations. The method is topological in nature, and gives the solution in terms of a global information gathering function[2] which determines the influence of input (boundary conditions) on output (the solution). In this work we will use this formalism to derive an explicit topological expression for the solution of a system of simultaneous linear equations.

In the discrete path approach, a homomorphism is established between the solution sought and a set of discrete paths. The paths originate at a number of sources which represent the initial conditions, and converge on a single vertex which represents the solution.[3] Furthermore, they are defined on a topological structure which is determined by the general nature of the problem, and are subjected to certain constraints reflecting its special particularities. Hence, the problem is essentially reduced to one of determining the underlying topological structure of the solution. For a system of simultaneous linear equations, this structure is the corresponding signal flow graph.[4]

Systems of simultaneous linear equations appear in every branch of theoretical physics, and it is certainly interesting to have a systematic topological method of handling them which is independent of Cramer's rule. When dealing with simple systems of linear equations, for which the appropriate determinants can be easily evaluated and simplified analytically, Cramer's rule is amply sufficient. On the other hand, for large systems of linear equations, the method presented here has the advantage of being transparent, of lending itself easily to a perturbative treatment,[5] and of being able to take full advantage of any symmetry that the system may have.[6] Cramer's rule has been given a graphic interpretation by Mason.[7] In a subsequent paper[8] we will give a completely topological derivation of Mason's theorem starting from the theorem derived in this work. For the purpose of comparison, an example will be worked out in the Appendix, using Cramer's rule, Mason's theorem, and the topological theorem derived here.

Faced with a solution of such a general character it is natural to question its practical usefulness in a given specific calculation. There is at least one important problem for which the details of the topological solution have been worked out. This is the problem of the radial Schrödinger equation with a linear central potential. All attempts to solve this problem in terms of special functions failed, and early workers turned to numerical[9] and approximate WKB solutions.[10] Using the method of Fröbenius,[11] the above Schrödinger equation can be transformed into an equivalent finite difference equation (an infinite system of simultaneous linear equations) which is then solved topologically, in accordance with the method presented here.[12] It is very significant from the point of view of the usefulness of the method to note that the topological expressions involved in the solution can subsequently be transformed into ordinary analytic expressions, essentially as sums over products of gamma functions.[13] This example provides tangible proof of the tractability of this type of topological solution, even when applied to a problem for which all other known methods[14] failed to give an analytic solution.

Notwithstanding the wide domain of applicability of the theorem derived here, from a particle physicist's point of view its most promising application is to the solution of the Schrödinger equation for a three-quark system (the baryon spectrum). The next step towards this goal would be the derivation, based on the theorem given here, of the solution for multivariable finite difference equations. This, in turn, will permit (using the techniques of Refs. 12 and 13) an analytic solution for those three-quark potentials that admit a multivariable power series expansion. The number of coefficients appearing in the expansion can be reduced considerably by making full use of the symmetry of the problem.[15] This constitutes a rather ambitions and long project, but at the same time each successive step in it is of inherent interest by itself.

II. TOPOLOGY OF A SYSTEM OF LINEAR EQUATIONS

A. Algebraic structure of "generalized paths"

Consider a graph $G(J,A)$, where J is the set of vertices and A the set of arcs. We refer to a sequence p of arcs belonging to A as a "generalized path." The length l_p of a path is the number of arcs in the sequence p. Let $\not{\Lambda} = \{p\}$ be the set of

[a)]Work supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/79/122375-05$01.00

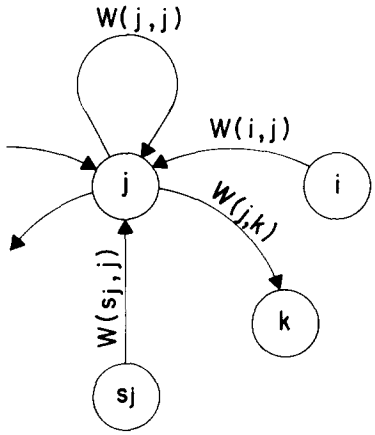

FIG. 1. A typical junction in a signal flow graph.

generalized paths and define the usual concatenation operation $\odot$ between two elements p_1 and p_2 of $\mathcal{P}$ by

$$p_1 \odot p_2 = (p_1, p_2) \in \mathcal{P}, \tag{2.1}$$

i.e., $p_1 \odot p_2$ is a sequence belonging to $\mathcal{P}$ obtained by taking the sequence p_1 followed by the sequence p_2. The identity element with respect to the operation $\odot$ is designated by e:

$$p \odot e = e \odot p = p . \tag{2.2}$$

e can be considered as a generalized path of length zero connecting a vertex to itself.

From the free monoid $(\mathcal{P}, \odot, e)$ we construct the power set $\mathcal{P}^*$ of $\mathcal{P}$ and introduce two operations $\cup$ (union) and $\otimes$ (concatenation) over $\mathcal{P}^*$. Let $P = \{p_1, p_2, \ldots, p_k\}$ and $P' = \{p'_1, p'_2, \ldots, p'_m\}$ be two elements of $\mathcal{P}^*$, then the operations $\cup$ and $\otimes$ are given by

$$P \cup P' = \{p_1, p_2, \ldots, p_k, p'_1, p'_2, \ldots, p'_m\} \in \mathcal{P}^* \tag{2.3a}$$

and

$$P \otimes P' = \{p_1 \odot p'_1, p_1 \odot p'_2, \ldots, p_1 \odot p'_m, \ldots, p_k \odot p'_m\} \in \mathcal{P}^* . \tag{2.3b}$$

The identity elements with respect to the $\cup$ and $\otimes$ operations are the empty set $\emptyset$ and the singleton $\{e\}$ respectively, both belonging to $\mathcal{P}^*$.

B. Signal flow graph

A system of N simultaneous linear equations can always be written in the form

$$x_j = \sum_{\substack{i \in \Gamma^{-1}(j) \\ i \neq s_j}} w(i,j) x_i + w(s_j, j), \quad j = 1,2,\ldots,N \tag{2.4}$$

and represented[4] by the signal flow graph G, as shown in Fig. 1, where $\Gamma^{-1}(j)$ is the set of vertices incident on vertex j, $w(i,j)$ is the weight of arc (i,j), and s_j is the source incident on vertex j. By definition, a vertex s_j is called a source, if there are no vertices incident on it.

For the signal flow graph we use the following notation, which is based on the algebraic structure introduced above:

$S = \{s_1, s_2, \cdots\}$ is the set of sources,
$p(n,j) \in \mathcal{P}$ = a path from vertex n to vertex j,
$P(n,j) \in \mathcal{P}^*$ = the set of paths from n to j,
$P^{(i)}(n,j) \in \mathcal{P}^*$ = the set of paths from n to j and passing through vertex i,
$P(j)$ = the set of paths from all sources to vertex j.

From the definition of $P(j)$ we have

$$P(j) = \cup_{s \in S} P(s,j) . \tag{2.5}$$

Furthermore, for two sources s_i and s_j,

$$P(s_i, s_j) = \begin{cases} \Phi, & i \neq j, \\ \{e\}, & i = j. \end{cases} \tag{2.6}$$

Thus, when j is a source, Eq. (2.5) reduces to

$$P(s) = \{e\} . \tag{2.7}$$

On the other hand, when j is not a source, the totality of paths from s to j can be decomposed as follows:

$$P(s,j) = \bigcup_{i \in \Gamma^{-1}(j)} P^{(i)}(s,j) . \tag{2.8}$$

Furthermore, for $i \in \Gamma^{-1}(j)$,

$$P^{(i)}(s,j) = P(s,i) \otimes \{(i,j)\} . \tag{2.9}$$

Combining Eqs. (2.5), (2.8), and (2.9), we have

$$P(j) = \cup_{s \in S} P(s,j) = \cup_{s \in S} \bigcup_{i \in \Gamma^{-1}(j)} P^{(i)}(s,j)$$

$$= \bigcup_{i \in \Gamma^{-1}(j)} \cup_{s \in S} P(s,i) \otimes \{i,j\}$$

or

$$P(j) = \bigcup_{i \in \Gamma^{-1}(j)} P(i) \otimes \{i,j\}, \quad \Gamma^{-1}(j) \neq \emptyset . \tag{2.10}$$

The above interesting topological relation defined on the signal flow graph G will be transformed into the set of linear equations (2.4) by an appropriate homomorphism defined below.

III. SOLUTION OF A SYSTEM OF LINEAR EQUATIONS

A. The homomorphism H

We establish a homomorphism from $(\mathcal{P}^*, \cup, \otimes)$ to $(R, +, \times)$ by defining the functional H, giving the value of an arc as being equal to its weight,

$$(\mathcal{P}^*, \cup, \otimes) \xrightarrow{H} (R, +, \times) , \tag{3.1a}$$

$$H(\{(i,j)\}) = w(i,j), \quad \{(i,j)\} \in \mathcal{P}^* . \tag{3.1b}$$

Since any element of $\mathcal{P}^*$ can be constructed from the arcs of the graph by using the operations of concatenation $\otimes$ and union $\cup$, Eqs. (3.1) completely define the value of any element of $\mathcal{P}^*$, i.e., the functional H. Naturally, we have

$$H(\emptyset) = 0, \quad H(\{e\}) = 1 , \tag{3.2}$$

Concretely, the above definition of H gives the value of a path as the product of the weight of its arcs, and the value of a set of paths as the sum of the values of its elements:

$$H(p) = \prod_{(i,j) \in p} w(i,j) , \tag{3.3a}$$

$$H(P) = \sum_{p \in P} H(p) . \tag{3.3b}$$

B. Condition of convergence

To study the convergence of the above formal expres-

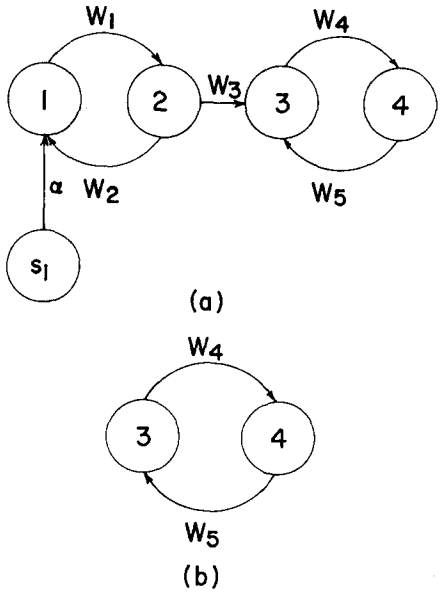

FIG. 2. (a) The signal flow graph G corresponding to the system of simultaneous equations (A1). (b) The signal flow graph $G(p_e(s_1,2))$.

sions for $H(P)$ in the case where P has an infinite number of elements, we will consider a complete graph with n vertices (excluding sources). Let $N_l(i,j)$ be the number of paths of length l from i to j, $P_l(i,j)$ the set of paths of length l from i to j, and w the maximum of the absolute values of the weights of the arcs of the graph. Then

$$P(i,j) = \bigcup_{l=1}^{\infty} P_l(i,j) \tag{3.4}$$

and

$$H(P_l(i,j)) \leqslant N_l(i,j) w^l ; \tag{3.5}$$

hence,

$$H(P(i,j)) \leqslant \sum_{l=1}^{\infty} N_l(i,j) w^l . \tag{3.6}$$

To evaluate $N_l(i,j)$ we make the following decomposition of the set $P_l(i,j)$:

$$P_l(i,j) = \bigcup_{k=1}^{n} P_{l-1}(i,k) \otimes \{(k,j)\} . \tag{3.7}$$

Thus,

$$N_l(i,j) = \sum_{k=1}^{n} N_{l-1}(i,k) . \tag{3.8a}$$

Furthermore,

$$N_1(i,j) = 1 . \tag{3.8b}$$

The solution of Eq. (3.8a) subject to the initial condition (3.8b) is easily seen to be

$$N_l(i,j) = n^{l-1} . \tag{3.9}$$

Replacing Eq. (3.9) in (3.6) we obtain

$$H(P(i,j)) \leqslant w \sum_{l=0}^{\infty} (nw)^l . \tag{3.10}$$

The series on the right hand side converges if and only if $nw < 1$.

C. The topological solution

Theorem (I): The value $H(P(j))$ of the set of all paths in the signal flow graph from all sources S to a vertex j is equal to x_j, the solution of the corresponding system of linear equations.

Proof: starting from the topological equation,

$$P(j) = \bigcup_{i \in \Gamma^{-1}(j)} P(j) \otimes \{(i,j)\}$$

isolating the contributions of the source s_j,

$$P(j) = \left[\bigcup_{\substack{i \in \Gamma^{-1}(j) \\ i \neq s_j}} P(i) \otimes \{(i,j)\} \right] \cup [P(s_j) \otimes \{(s_j,j)\}]$$

and making use of Eq. (2.7), we obtain

$$P(j) = \left[\bigcup_{\substack{i \in \Gamma^{-1}(j) \\ i \neq s_j}} P(i) \otimes \{(i,j)\} \right] \cup [\{e\} \otimes \{(s_j,j)\}] . \tag{3.11}$$

The rest of the proof consists of applying the homomorphism H to Eq. (3.11) to reproduce the system of linear equations(2.1):

$$H(P(j)) = \sum_{\substack{i \in \Gamma^{-1}(j) \\ i \neq s_j}} H(P(i)) H(\{(i,j)\}) + H(\{e\}) H(\{(s_j,j)\}) .$$

Making use of Eqs. (3.1) and (3.2), the above equation reduces to

$$H(P(j)) = \sum_{\substack{i \in \Gamma^{-1}(j) \\ i \neq s_j}} H(P(i)) w(i,j) + w(s_j,j) . \tag{3.12}$$

Comparing Eqs. (3.5) and (2.1) shows that $H(P(j))$ is a solution of the system of linear equations (2.1).

An example illustrating the application of the above theorem is worked out in the Appendix.

IV. CONCLUSION

We have shown that the solution of a system of simultaneous linear equations is the value of a given set of paths on the corresponding signal flow graph. The value of this type of solution and its superiority to Cramer's rule is most evi-

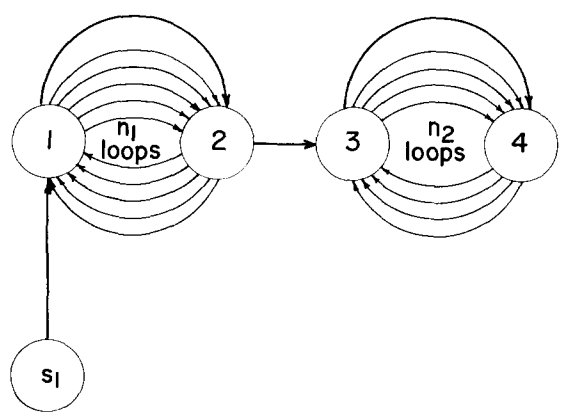

FIG. 3. A typical path from source s_1 to vertex 4 on the signal flow graph G corresponding to the system of simultaneous equations (A1).

TABLE I. Evaluating the determinants of the graph G of Fig. 2 and its subgraphs $G(p_e(s_1,2))$ and $G(p_e(s_1,4))$.

Graph	L_1	L_2	$L_k (k\geqslant 3)$	Δ
G	$\lambda_1+\lambda_2$	$\lambda_1\lambda_2$	0	$1-(\lambda_1+\lambda_2)+\lambda_1\lambda_2$
$G(p_e(s_1,2))$	λ_2	0	0	$1-\lambda_2$
$G(p_e(s_1,4))$	0	0	0	1

dent for large systems of equations with a high degree of internal symmetry. Nevertheless, as can be seen from the example worked out in the Appendix, even for a simple system of simultaneous linear equations, the solution presented here is simpler and faster than that obtained by Cramer's rule or by Mason's theorem. Furthermore, it gives the answer in a form that is already simplified, factorized, and written in terms of the natural parameters of the problem.

APPENDIX: AN ILLUSTRATIVE APPLICATION

The signal flow graph shown in Fig. 2(a) represents the following system of equations:

$$\begin{aligned} x_1 &= w_2x_3+\alpha\,,\\ x_2 &= w_1x_1\,,\\ x_3 &= w_3x_2+w_5x_4\,,\\ x_4 &= w_4x_3\,. \end{aligned} \tag{A1}$$

Let

$$\lambda_1 = w_1w_2\,,\quad \lambda_2 = w_4w_5\,. \tag{A2}$$

As an example we will compute x_2 and x_4 using successively theorem 1, Mason's theorem, and Cramer's rule.

1. Theorem 1

Consider a typical path shown in Fig. 3 from source s_1 to vertex 4. The value of this path is $\alpha\lambda_1^{n_1}w_1w_3w_4\lambda_2^{n_2}$, where n_1 and n_2 are nonnegative integers. To every pair of values (n_1,n_2) corresponds a path belonging to $P(4)$. Hence,

$$\begin{aligned} x_4 = H(P(4)) &= \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\alpha\lambda_1^{n_1}w_1w_3w_4\lambda_2^{n_2}\\ &= \alpha w_1w_3w_4\sum_{n_1=0}^{\infty}\lambda_1^{n_1}\sum_{n_2=0}^{\infty}\lambda_2^{n_2}\,. \end{aligned}$$

If $|\lambda_1|<1$ and $|\lambda_2|<1$, we have

$$x_4 = \alpha w_1w_3w_4\left(\frac{1}{1-\lambda_1}\right)\left(\frac{1}{1-\lambda_2}\right). \tag{A3a}$$

In a similar fashion we find that

$$x_2 = H(P(2)) = \alpha w_1\sum_{n=0}^{\infty}\lambda_1^{n}$$

or

$$x_2 = \frac{\alpha w_1}{1-\lambda_1}\,. \tag{A3b}$$

2. Mason's theorem

According to Mason's theorem,

$$x_i = \frac{\Sigma_{s\in S}\Sigma_{p_e(s,i)}H(p_e(s,i))\Delta(G(p_e(s,i)))}{\Delta(G)}\,, \tag{A4}$$

where $\Delta(G)$ is the "determinant" of the signal flow graph:

$$\Delta(G) = 1-L_1+L_2-L_3+\cdots+(-1)^kL_k\,, \tag{A5}$$

L_k being the sum of the products of the values of k disjoint elementary circuits in the signal flow graph. $p_e(s,i)$ is an elementary path from the source s to the vertex i, and $G(p_e(s,i))$ is the subgraph obtained from G by eliminating all the vertices of $p_e(s,i)$.

For the flow graph under consideration, there is only one elementary path, $p_e(s_1,2)=(s_1,1,2)$, from s_1 to vertex 2, and only one elementary path, $p_e(s_1,4)=(s_1,1,2,3,4)$, from s_1 to vertex 4. The values of these paths are given by

$$H(p_e(s_1,2)) = \alpha w_1 \quad\text{and}\quad H(p_e(s_1,4)) = \alpha w_1w_3w_4\,. \tag{A6}$$

Corresponding to the above two elementary paths, there are two subgraphs; $G(p_e(s_1,4))$ which is empty, and $G(p_e(s_1,2))$ shown in Fig. 2(b). The determinants for these two subgraphs, as well as the determinant for graph G, are evaluated in Table I. Thus using the above results we find that the application of Mason's theorem to calculate x_2 and x_4 gives:

$$x_4 = \frac{\alpha w_1w_3w_4}{1-(\lambda_1+\lambda_2)+\lambda_1\lambda_2}\,, \tag{A7a}$$

$$x_2 = \frac{\alpha w_1(1-\lambda_2)}{1-(\lambda_1+\lambda_2)+\lambda_1\lambda_2}\,. \tag{A7b}$$

3. Cramer's rule

Finally, according to Cramer's rule

$$x_i = N_i/D\,, \tag{A8}$$

where

$$\begin{aligned} D &= \begin{vmatrix} 1 & -w_2 & 0 & 0\\ -w_1 & 1 & 0 & 0\\ 0 & -w_3 & 1 & -w_5\\ 0 & 0 & -w_4 & 1 \end{vmatrix}\\ &= (1-w_1w_2)(1-w_4w_5)\,, \end{aligned} \tag{A9}$$

$$\begin{aligned} N_4 &= \begin{vmatrix} 1 & -w_2 & 0 & \alpha\\ -w_1 & 1 & 0 & 0\\ 0 & -w_3 & 1 & 0\\ 0 & 0 & -w_4 & 0 \end{vmatrix}\\ &= \alpha w_1w_3w_4\,, \end{aligned} \tag{A10a}$$

$$\begin{aligned} N_2 &= \begin{vmatrix} 1 & \alpha & 0 & 0\\ -w_1 & 0 & 0 & 0\\ 0 & 0 & 1 & -w_5\\ 0 & 0 & -w_4 & 1 \end{vmatrix}\\ &= \alpha w_1(1-w_4w_5)\,. \end{aligned} \tag{A10b}$$

Hence,

$$x_4 = \frac{\alpha w_1 w_3 w_4}{(1 - w_1 w_2)(1 - w_4 w_5)}, \tag{A11a}$$

$$x_2 = \frac{\alpha w_1}{(1 - w_1 w_2)}. \tag{A11b}$$

It is easy to see that Eqs. (A3a), (A7a), and (A11a) are all equivalent, and that Eqs. (A3b), (A7b), and (A11b) are also equivalent.

[1]A.F. Antippa, J. Math. Phys. **18**, 2214 (1977).

[2]This is essentially the "combinatorics function" of Ref. 1. In engineering terms it is a signal flow function.

[3]In conformity with the language of graph theory the direction of the paths is taken to be opposite to that used in Ref. 1, i.e., "terminal points" in Ref. 1 are "sources" in this work.

[4]W.K. Chen, *Applied Graph Theory* (North-Holland, Amsterdam, 1971); W. Mayeda, *Graph Theory* (Wiley Interscience, New York, 1972).

[5]The importance of the contribution of a given path to the solution is inversely proportional to its length.

[6]An example in point is the special case of finite difference equations. These are an infinite system of simultaneous linear equations with a very high degree of internal symmetry. Cramer's rule is practically useless in this case (see Ref. 14), while the method presented here can handle these equations with relative ease (see Ref. 1).

[7]S.J. Mason, Proc. I.R.E. **45**, 829 (1957).

[8]N.K. Toan and A.F. Antippa, "Derivation of Mason's theorem by the path formalism" (to be published).

[9]E.Eichten, K. Gottfried, T. Kinushita, J. Kogut, and T.M. Yan, Phys. Rev. Lett. **34**, 370 (1975); J. Gunion and R. Willey, Phys. Rev. D **12**, 174 (1975).

[10]B. Harrington, S. Park, and A. Yildiz, Phys. Rev. Lett. **34**, 168 (1975); **34**, 704 (1975); H. Muller-Kirsten, Phys. Rev. D **12**, 1103 (1975).

[11]H. Jeffreys and B.S. Jeffreys, *Methods of Mathematical Physics* (Cambridge University, Cambridge, England, 1972).

[12]A.F. Antippa and A.J. Phares, J. Math. Phys. **18**, 173 (1977); **19**, 308 (1978).

[13]A.F. Antippa and N.K. Toan, Can. J. Phys. **57**, 417 (1979).

[14]P.M. Morse and H. Feshback, *Methods of Mathematical Physics* (McGraw-Hill, New York, 1953), Vol. 1, pp. 568–576.

[15]J. Patera, R.T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978).

Ladder operators of group matrix elements

C. K. E. Schneider[a)]
Institut für Theorie der Elementarteilchen, Freie Universität Berlin, Arnimallee 3, 1 Berlin 33, Germany

Raj Wilson
Max-Planck Institut, Riemerschmidstr. 7, D-8130 Starnberg, Germany
and Department of Theoretical Physics, University of Madras, Madras 600025, India

(Received 7 July 1978)

All ladder operators and some recurrence relations of the matrix elements of certain group elements of SO(3), SO(2,1), E(2), SO(4), SO(3,1), and E(3) have been explicitly determined and the underlying factorizations of the second- and the fourth-order linear ordinary differential equations in terms of first- and second-order ladder operators have been transparently demonstrated as an extension to the Schrödinger–Infeld–Miller factorization. These ladder operators are very useful in physical applications where the corresponding matrix elements represent certain physical transitions.

I. INTRODUCTION AND SUMMARY

Schrödinger proposed a simple method of factorization[1] of a quantum-mechanical second-order linear differential equation into a product of two first-order differential operators (ladder operators or stair operators). Each ladder operator has a special property that when it acts on the original eigenfunction, a new eigenfunction is created with a quantum number raised or lowered by one unit. Schrödinger's method has been further systematically extended to a class of second-order linear differential equations and to their solutions by Infeld and Inui[1] and various recurrence relations satisfied by the solutions have been derived.

A systematic group theoretical treatment to Schrödinger–Infeld's analytic factorization is given by Miller.[2] Here, the second-order linear differential equation appears as the invariant Casimir product, in the Lie algebra of a Lie group of rank one, say SO(3), acting on the basis function belonging to a given unitary irreducible representation of the group. The factorization is obtained through the observation that the Casimir product can be expressed in terms of products of two group generators which can be suitably chosen to be the first-order ladder operators. Recently, Miller's method has been further applied in determining first-order ladder operators admitted by certain type of Hamiltonians in two variables.[3]

The analytic and group theoretic factorization procedures mentioned above have been so far restricted only to second-order differential equations and hence in obtaining the first-order ladder operators within the group (say, G) of rank one. Nevertheless, by going outside G, for example to the direct product group $G \otimes G$, we can obtain additional factorizations of the same differential equation. More precisely, this amounts to algebraically constructing ladder operators that raise and lower the Casimir labels in G. Thus for orthogonal groups the number of additional factorizations we could achieve by going to $G \otimes G$ is equal to the rank of the Group G. It is the rank minus one for the Euclidean groups, provided we consider their unitary irreducible representations by reducing with respect to their maximal orthogonal subgroups. Furthermore, the factorizations of higher-order differential equations can be seen by the construction of higher-order ladder operators in larger groups. For example, the fourth order differential equation for SO(4) can be factorized in terms of second-order ladder operators.

The motivation of this paper is therefore to expound and advocate the above mentioned constructions, as an extension to the Schrödinger–Infeld–Miller factorization, using the orthogonal and Euclidean groups of three and six parameters. Our approach is to some degree different from that of Miller. While, in the case of Miller, the ladder operators act in the space of eigenfunctions, our ladder operators are constructed to act in a space of certain transition matrix elements. Although mathematically both approaches are equivalent as in both cases we deal with the same kind of special functions, physically we seem to gain certain advantages. Since the transition matrix elements or transition probabilities are related to certain physical observables, our ladder operators obviously acquire physical significance. They transform, in a trivial way, a given matrix element to the one in which the concerned quantum number in one of the eigenstates is raised or lowered by one unit. Thus they relate two different physical situations in a simple way (for example, the elastic and inelastic form factors). Furthermore, once the complete set of independent ladder operators of a given matrix element is determined, then not only do we know the different factorizations of the differential equations involved but also in principle we can obtain all recurrence relations satisfied by the matrix element. The latter ensures practical advantages.

The material of this paper is arranged as follows: In Sec. II we deal with the groups SO(3), SO(2,1), and E(2), determine explicitly certain recurrence relations and all ladder

[a)]Supported partially by Deutscher Akademischer Austauschdienst.

 0022-2488/79/122380-11$01.00

operators of Wigner, Bargmann, and Inönü–Wigner functions and demonstrate the different ways of factorizations. In Sec. III we give the same treatment as in Sec. II for the groups SO(4), SO(3,1), and E(3), define Dolginov–Biedenharn, Dolginov–Toptygin, and Vilenkin–Akim–Levin functions, and eventually show how a fourth-order differential equation can be factorized in terms of second-order ladder operators.

II. THE SO(3), SO(2,1), AND E(2) MATRIX ELEMENTS

We first consider the generators L_i, let $i = 1,2,3$; and the complete orthogonal basis states of SO(3), Φ_{lm}, let $l = 0,1,\ldots,\infty$, $-l \leqslant m \leqslant l$. The Wigner rotation function, defined by the matrix element

$$(\Phi_{lm'}, \exp(-i\theta L_2)\Phi_{lm}) \equiv d^{l}_{m'm}(\theta) \tag{II.1}$$

has an explicit integral form[4] (c: unit circle)

$$d^{l}_{m'm}(\theta) = \frac{i^{m'-m}}{2\pi i}\left(\frac{(l-m')!(l+m')!}{(l-m)!(l+m)!}\right)^{1/2}\oint_c dx\, x^{m'-l-1} \times\left(x\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}\right)^{l-m}\left(\cos\frac{\theta}{2} + ix\sin\frac{\theta}{2}\right)^{l+m}, \tag{II.2}$$

as of a hypergeometric function. If one differentiates (II.1) with respect to the group parameter θ, using the identity (Baker–Campbell–Hausdorff formula)

$$\exp(\pm i\theta L_2)L_3\exp(\mp i\theta L_2) = L_3\cos\theta \mp L_1\sin\theta,$$

and the actions of the generators L_3 and L_1 on the basis states Φ_{lm} one easily obtains the following familiar recurrence relations:

$$\begin{aligned}
&\frac{d}{d\theta}d^{l}_{m'm}(\theta) = -\tfrac{1}{2}[(l-m)(l+m+1)]^{1/2}d^{l}_{m'\,m+1}(\theta) + \tfrac{1}{2}[(l+m)(l-m+1)]^{1/2}d^{l}_{m'\,m-1}(\theta),\\
&\frac{d}{d\theta}d^{l}_{m'm}(\theta) = \tfrac{1}{2}[(l-m')(l+m'+1)]^{1/2}d^{l}_{m'+1\,m}(\theta) - \tfrac{1}{2}[(l+m')(l-m'+1)]^{1/2}d^{l}_{m'-1\,m}(\theta),\\
&\left(\frac{m'-m\cos\theta}{\sin\theta}\right)d^{l}_{m'm}(\theta) = -\tfrac{1}{2}[(l-m)(l+m+1)]^{1/2}d^{l}_{m'\,m+1}(\theta) - \tfrac{1}{2}[(l+m)(l-m+1)]^{1/2}d^{l}_{m'\,m-1}(\theta),\\
&\left(\frac{m-m'\cos\theta}{\sin\theta}\right)d^{l}_{m'm}(\theta) = \tfrac{1}{2}[(l-m')(l+m'+1)]^{1/2}d^{l}_{m'+1\,m}(\theta) + \tfrac{1}{2}[(l+m')(l-m'+1)]^{1/2}d^{l}_{m'-1\,m}(\theta)
\end{aligned} \tag{II.3}$$

Solving (II.3), the ladder operators of $d^{l}_{m'm}(\theta)$ with respect to m and m' are obtained:

$$\begin{aligned}
&\mp[(l\mp m)(l\pm m+1)]^{-1/2}\left(\frac{d}{d\theta} \pm \frac{m'-m\cos\theta}{\sin\theta}\right)d^{l}_{m'm}(\theta) = L^{\pm}_{m}d^{l}_{m'm}(\theta) = d^{l}_{m'\,m\pm1}(\theta),\\
&\pm[(l\mp m')(l\pm m'+1)]^{-1/2}\left(\frac{d}{d\theta} \pm \frac{m-m'\cos\theta}{\sin\theta}\right)d^{l}_{m'm}(\theta) = L^{\pm}_{m'}d^{l}_{m'm}(\theta) = d^{l}_{m'\pm1\,m}(\theta)
\end{aligned} \tag{II.4}$$

Indeed, these ladder operators for SO(3) are well known.

In order to obtain the corresponding ladder operators which raise and lower the Casimir parameter l in $d^{l}_{m'm}(\theta)$ we have to go outside SO(3) and consider, for example, the representations of the direct product group SO(3) ⊗ SO(3). In this representation space, the Wigner functions satisfy the Clebsh–Gordan series,

$$d^{l_1}_{m'_1m_1}(\theta)d^{l_2}_{m'_2m_2}(\theta) = \sum_{l}(2l+1)\begin{pmatrix} l_1 & l_2 & l\\ m'_1 & m'_2 & m'\end{pmatrix}\begin{pmatrix} l_1 & l_2 & l\\ m_1 & m_2 & m\end{pmatrix}d^{l}_{m'm}(\theta), \tag{II.5}$$

$\begin{pmatrix}\cdots\end{pmatrix}$ being the 3-j symbol. Evaluating (II.5) for special values: $l_2 = 1, m'_2 = m_2 = 0$; $l_2 = 1, m_2 = 1, m'_2 = 0$, and $l_2 = 1, m'_2 = 1$, $m_2 = 0$, we obtain the recurrence relations:

$$\begin{aligned}
(2l+1)[l(l+1)\cos\theta - mm']d^{l}_{m'm}(\theta) &= (l+1)[(l-m')(l+m')(l-m)(l+m)]^{1/2}d^{l-1}_{m'm}(\theta)\\
&\quad + l[(l-m'+1)(l+m'+1)(l-m+1)(l+m+1)]^{1/2}d^{l+1}_{m'm}(\theta),\\
(2l+1)l(l+1)\sin\theta d^{l}_{m'm}(\theta) &= -(l+1)[(l-m')(l+m')(l-m)(l-m-1)]^{1/2}d^{l-1}_{m'\,m+1}(\theta)\\
&\quad -(2l+1)[(l-m)(l+m+1)]^{1/2}m'd^{l}_{m'\,m+1}(\theta)\\
&\quad + l[(l-m'+1)(l+m'+1)(l+m+1)(l+m+2)]^{1/2}d^{l+1}_{m'\,m+1}(\theta),
\end{aligned} \tag{II.6}$$

$$(2l+1)l(l+1)\sin\theta d^{l}_{m'm}(\theta) = (l+1)[(l-m')(l-m'-1)(l-m)(l+m)]^{1/2} d^{l-1}_{m'+1\ m}(\theta)$$

$$+ (2l+1)[(l-m')(l+m'+1)]^{1/2} m d^{l}_{m'+1\ m}(\theta)$$

$$- l[(l+m'+1)(l+m'+2)(l-m+1)(l+m+1)]^{1/2} d^{l+1}_{m'+1\ m}(\theta).$$

The last two relations are related by $m' \leftrightarrow m$ and $d^{l}_{m'm}(\theta) = (-1)^{m'-m} d^{l}_{mm'}(\theta)$. From (II.6) and (II.4) we obtain the ladder operators:

$$-l[(l-m')(l+m')(l-m)(l+m)]^{-1/2}\left(\sin\theta\frac{d}{d\theta} - l\cos\theta + \frac{mm'}{l}\right) d^{l}_{m'm}(\theta) = L_{l}^{-} d^{l}_{m'm}(\theta) = d^{l-1}_{m'm}(\theta),$$

$$(l+1)[(l-m'+1)(l+m'+1)(l-m+1)(l+m+1)]^{-1/2}\left(\sin\theta\frac{d}{d\theta} + (l+1)\cos\theta - \frac{mm'}{(l+1)}\right) d^{l}_{m'm}(\theta)$$

$$= L_{l}^{+} d^{l}_{m'm}(\theta) = d^{l+1}_{m'm}(\theta). \tag{II.7}$$

The operators $L_{l}^{\pm}$ are related by the substitution $l \leftrightarrow -l-1$ which keeps the eigenvalue $l(l+1)$ of the SO(3)-Casimir product unchanged.

Thus, we have obtained all the four ladder operators $L_{m}^{\pm}$ ($L_{m'}^{\pm}$) and $L_{l}^{\pm}$ and these operators satisfy the operator equation

$$[L_{r\pm1}^{\mp} L_{r}^{\pm} - 1] d^{l}_{m'm}(\theta) = 0, \quad r = (m,l), \tag{II.8}$$

which in turn is equivalent to the linear second ordinary differential equation

$$\left(\frac{d^2}{d\theta^2} + \cot\theta\frac{d}{d\theta} - \frac{1}{\sin^2\theta}(m'^2 + m^2 - 2m'm\cos\theta) + l(l+1)\right) d^{l}_{m'm}(\theta) = 0. \tag{II.9}$$

The solution to this equation, finite for $\cos\theta = \pm 1$ is given by (II.2). Clearly, (II.8) indicates that (II.9) can be factorized in two different ways and the factorizations of "Type A" ($r = m$) and "Type E" ($r = l$) in the sense of Infeld and Hull.[1] They differ with respect to the group parameter θ, by an overall multiplication by $\sin\theta$.

Next, we consider the generators L_i, $i = 1,2,3$ of SO(2,1). They satisfy the Lie commutations:

$$[L_1,L_2] = -iL_3, \quad [L_1,L_3] = -iL_2, \quad [L_2,L_3] = iL_1.$$

In a given unitary irreducible representation of SO(2,1)[5] we choose a standard orthonormal discrete basis Φ_{Kn} such that the Casimir product

$$(L_3^2 - L_1^2 - L_2^2)\Phi_{Kn} = K(K+1)\Phi_{Kn} \quad \text{and} \quad L_3\Phi_{Kn} = n\Phi_{Kn}.$$

For the discrete principal series D_{K}^{+} (D_{K}^{-}), $-K = \frac{1}{2}, 1, \frac{3}{2}, \cdots; n = -K, -K+1, \cdots$ ($n = K, K-1, \cdots$). The matrix element

$$(\Phi_{Kn'}, \exp(-i\theta L_2)\Phi_{Kn}) \equiv V^{K}_{n'n}(\theta) \tag{II.10}$$

defines the Bargmann function which can be expressed in an integral form as

$$V^{K}_{n'n}(\theta) = \frac{i^{n'-n}}{2\pi i}\left(\frac{(K-n')!(K+n')!}{(K-n)!(K+n)!}\right)^{1/2} \oint_{c} dx\, x^{n'-K-1}\left(x\cosh\frac{\theta}{2} - \sinh\frac{\theta}{2}\right)^{K-n}\left(\cosh\frac{\theta}{2} - x\sinh\frac{\theta}{2}\right)^{K+n}. \tag{II.11}$$

In order to obtain the ladder operators for SO(2,1), one may directly follow the procedure adopted for SO(3). In this case, exact expressions for the SO(2,1) Clebsch–Gordan coefficients[6] must be used. However, it is much easier to deduce the expressions for SO(2,1) from those for SO(3) by means of the analytic continuation $\theta \to i\theta$ (Weyl's unitarian trick) with the formal substitutions $l = K$, $m' = n'$, $m = n$. We give below the four ladder operators:

$$\pm i[(K \mp n)(K \pm n + 1)]^{-1/2}\left(\frac{d}{d\theta} \pm \frac{n' - n\cosh\theta}{\sinh\theta}\right) V^{K}_{n'n}(\theta) = L_{n}^{\pm} V^{K}_{n'n}(\theta) = V^{K}_{n'\ n\pm1}(\theta),$$

$$\mp i[(K \mp n')(K \pm n' + 1)]^{-1/2}\left(\frac{d}{d\theta} \pm \frac{n - n'\cosh\theta}{\sinh\theta}\right) V^{K}_{n'n}(\theta) = L_{n'}^{\pm} V^{K}_{n'n}(\theta) = V^{K}_{n'\pm1\ n}(\theta),$$

$$-K[(K-n')(K+n')(K-n)(K+n)]^{-1/2}\left(\sinh\theta\frac{d}{d\theta} - K\cosh\theta + \frac{n'n}{K}\right) V^{K}_{n'n}(\theta) = L_{K}^{-} V^{K}_{n'n}(\theta) = V^{K-1}_{n'n}(\theta),$$

$$(K+1)[(K-n'+1)(K+n'+1)(K-n+1)(K+n+1)]^{-1/2}\left(\sinh\theta\frac{d}{d\theta} + (K+1)\cosh\theta - \frac{n'n}{(K+1)}\right) V^{K}_{n'n}(\theta)$$

$$= L_{K}^{+} V^{K}_{n'n}(\theta) = V^{K+1}_{n'n}(\theta). \tag{II.12}$$

These ladder operators satisfy the operator equations [as in SO(3)],

$$[L_{r\pm1}^{\mp}L_r^{\pm}-1]V_{n'n}^K(\theta)=0,\quad r=(n,K), \tag{II.13}$$

which display the two "Type A" ($r=n$) and "Type E" ($r=K$) factorizations of the second order,

$$\left(\frac{d^2}{d\theta^2}+\coth\theta\frac{d}{d\theta}-\frac{1}{\sinh\theta}(n'^2+n^2-2n'n\cosh\theta)-K(K+1)\right)V_{n'n}^K(\theta)=0. \tag{II.14}$$

The solution of this equation, continuous at the point $\cosh\theta=1$, is given by (II.11). Again, as in SO(3), the two factorizations differ mainly by a multiplication factor of $\sinh\theta$.

It should be noted that since Wigner and Bargmann functions are given essentially by the ${}_2F_1$ hypergeometric functions, all these ladder operators can be readily obtained by using Gauss' contiguous relations.

Next, we consider the generators of E(2)—the two-dimensional Euclidean group—L_i, let $i=1,2,3$. They satisfy the commutations

$$[L_1,L_2]=0,\quad [L_1,L_3]=-iL_2,\quad [L_2,L_3]=iL_1.$$

The faithful unitary irreducible representations of E(2) may be realized over discrete basis states $\Phi_{\tau m}$, $0\leqslant\tau^2<\infty$, $m=0,\pm1,\cdots$ such that the Casimir product

$$(L_1^2+L_2^2)\Phi_{\tau m}=\tau^2\Phi_{\tau m}\quad\text{and}\quad L_3\Phi_{\tau m}=m\Phi_{\tau m}.$$

The matrix element

$$(\Phi_{\tau m'},\exp(-iyL_2)\Phi_{\tau m})\equiv J_{m'm}(\tau y), \tag{II.15}$$

defines the Inönü–Wigner function[7] which has a simple integral form

$$J_{m'm}(\tau y)=(-1)^{m'-m}J_{m'-m}(\tau y)$$

$$=\frac{(-1)^{m'-m}}{2\pi i}\oint_c dx\,x^{m-m'-1}\exp\left[\tfrac{1}{2}\tau y\left(x-\frac{1}{x}\right)\right], \tag{II.16}$$

of the generalized Bessel function of integral order. Now, as in the case for SO(3), we differentiate (II.15) with respect to the group parameter y (translation), use the identity

$$\exp(\pm iyL_2)L_3\exp(\mp iyL_2)=L_3\mp yL_1,$$

and obtain the following recurrence relations:

$$\begin{aligned}
y\frac{d}{dy}J_{m'm}(\tau y)&=-\tfrac{1}{2}\tau yJ_{m'\,m+1}(\tau y)+\tfrac{1}{2}\tau yJ_{m'\,m-1}(\tau y),\\
y\frac{d}{dy}J_{m'm}(\tau y)&=\tfrac{1}{2}\tau yJ_{m'+1\,m}(\tau y)-\tfrac{1}{2}\tau yJ_{m'-1\,m}(\tau y),\\
(m'-m)J_{m'm}(\tau y)&=-\tfrac{1}{2}\tau yJ_{m'\,m+1}(\tau y)-\tfrac{1}{2}\tau yJ_{m'\,m-1}(\tau y),\\
(m-m')J_{m'm}(\tau y)&=\tfrac{1}{2}\tau yJ_{m'+1\,m}(\tau y)+\tfrac{1}{2}\tau yJ_{m'-1\,m}(\tau y).
\end{aligned} \tag{II.17}$$

From these relations we deduce the ladder operators of $J_{m'm}(\tau y)$ with respect to m', m:

$$\begin{aligned}
\mp\frac{1}{\tau y}\left(y\frac{d}{dy}\pm(m'-m)\right)J_{m'm}(\tau y)&=L_m^{\pm}J_{m'm}(\tau y)=J_{m'\,m\pm1}(\tau y),\\
\pm\frac{1}{\tau y}\left(y\frac{d}{dy}\pm(m-m')\right)J_{m'm}(\tau y)&=L_{m'}^{\pm}J_{m'm}(\tau y)=J_{m'\pm1\,m}(\tau y)
\end{aligned} \tag{II.18}$$

These operators satisfy the operator equation

$$[L_{m\pm1}^{\mp}L_m^{\pm}-\mathbb{1}]J_{m'm}(\tau y)=0, \tag{II.19}$$

which shows the single "Type C" factorization for the second-order differential equation

$$\left(y^2\frac{d^2}{dy^2}+y\frac{d}{dy}-(m'-m)^2+\tau^2y^2\right)J_{m'm}(\tau y)=0. \tag{II.20}$$

The solution to this equation, finite for $y=0$ is given by (II.16).

We now define three generators: $L_1'=\sqrt{\epsilon}L_1$, $L_2'=\sqrt{\epsilon}L_2$, $L_3'=L_3$; $\epsilon=+1,-1,0$, of a "master group" $G(3,\epsilon)$ such that L_i and L_i' satisfy the Lie commutations:

$$[L_i,L_j]=i\epsilon_{ijk}L_k,\quad [L_1',L_2']=\epsilon iL_3',\quad [L_1',L_2']=-iL_2',\quad [L_2',L_3']=iL_1'.$$

It is now transparent that $G(3;+1)=\mathrm{SO}(3)$, $G(3;-1)=\mathrm{SO}(2,1)$ and in the limit $\epsilon\to0$ (Inönü–Wigner contraction)

$G(3;0) = E(2)$. The Casimir product of $G(3;\epsilon)$ is given by

$$\epsilon L_1^2 + \epsilon L_2^2 + L_3^2 = \alpha(\alpha+1)\mathbb{1},$$

where for $\epsilon = +1$, $\alpha = l$; $\epsilon = -1$, $\alpha = K$ and in the limit $\epsilon \to 0$ and $\alpha \to \infty$; $\epsilon\alpha^2 = \tau^2$. Thus, one can construct four "master" ladder operators in terms of $\sqrt{\epsilon}$ from which those for SO(3), SO(2,1), and E(2) can be appropriately deduced as special and limiting cases. Such operators are trivially obtained from those of SO(3) with the substitution $\theta \to \theta\sqrt{\epsilon}$ and $l = \alpha$. For E(2) we take $\theta \to y$ and the limit $\epsilon \to 0$ and $\alpha \to \infty$ with $\alpha\sqrt{\epsilon} = \tau$. Under this limit the integral form (II.2) goes over to (II.16) in virtue of the Mehler–Heine asymptotic formula of Jacobi polynomials.

III. THE SO(4), SO(3,1), AND E(3) MATRIX ELEMENTS

The six generators L_i, M_i, $i = 1,2,3$, of SO(4) satisfy the Lie-commutations

$$[L_i,L_j] = i\epsilon_{ijk}L_k, \quad [M_i,M_j] = i\epsilon_{ijk}M_k, \quad [L_i,M_j] = i\epsilon_{ijk}M_k.$$

In a unitary irreducible representation of SO(4) we choose the canonical orthonormal basis states $\Phi^{nl_0}_{lm}$, $l_0 = 0, \pm\frac{1}{2}, \pm 1,\ldots$, $n = 1 + |l_0|,\ldots,\infty$, $l = |l_0|, 1+|l_0|,\ldots,n-1$, such that the two Casimir products

$$(\mathbf{L}^2 + \mathbf{M}^2)\Phi^{nl_0}_{lm} = (n^2 + l_0^2 - 1)\Phi^{nl_0}_{lm}, \quad (\mathbf{L}\cdot\mathbf{M})\Phi^{nl_0}_{lm} = nl_0\Phi^{nl_0}_{lm}, \tag{III.1}$$

and l, m are the usual eigenvalues belonging to the SO(3) subgroup.

The Dolginov–Biedenharn function[8] is defined by the matrix elements

$$(\Phi^{nl_0}_{l'm}, \exp(-i\theta M_3)\Phi^{nl_0}_{lm}) \equiv D^{[nl_0]}_{l'ml}(\theta), \tag{III.2}$$

which has a simple integral form

$$D^{[nl_0]}_{l'ml}(\theta) = \tfrac{1}{2}[(2l'+1)(2l+1)]^{1/2}\left(\frac{(n+l')!(n-l'-1)!}{(n+l)!(n-l-1)!}\right)^{1/2}$$

$$\times \int_{-1}^{+1} dx\,(\cos\theta - ix\sin\theta)^{n-1} d^{l'*}_{l_0m}(x) d^{l}_{l_0m}\left(\frac{x\cos\theta - i\sin\theta}{\cos\theta - ix\sin\theta}\right). \tag{III.3}$$

Furthermore, the local isomorphism $SO(4) \sim SO(3) \otimes SO(3)$ enables one to express the matrix elements in terms of 3-j symbols as,

$$D^{[nl_0]}_{l'ml}(\theta) = [(2l'+1)(2l+1)]^{1/2} \sum_{m_1m_2} \begin{pmatrix} l_+ & l_- & l' \\ m_1 & m_2 & m \end{pmatrix}\begin{pmatrix} l_+ & l_- & l \\ m_1 & m_2 & m \end{pmatrix} \exp[-i\theta(m_1 - m_2)],$$

$$l_\pm = \tfrac{1}{2}(n-1) \pm \tfrac{1}{2}l_0, \qquad m = m_1 + m_2. \tag{III.4}$$

We now differentiate (III.2) with respect to the group parameter, use the identity,

$$\exp(i\theta M_3)H_\pm \exp(-i\theta M_3) = H_\pm\cos\theta \pm F_\pm\sin\theta, \quad H_\pm = L_1 \pm iL_2, \quad F_\pm = i(M_1 \pm iM_2), \tag{III.5}$$

and the actions[9] of the operators M_3, $H_\pm$, and $F_\pm$ on the basis states $\Phi^{nl_0}_{lm}$ and obtain the following recurrence relations:

$$\left(\frac{d}{d\theta} + i\frac{ml_0n}{l(l+1)}\right)D^{[nl_0]}_{l'ml}(\theta)$$

$$= \frac{i}{l}\left(\frac{(l-l_0)(l+l_0)(n-l)(n+l)(l-m)(l+m)}{(2l-1)(2l+1)}\right)^{1/2} D^{[nl_0]}_{l'ml-1}(\theta)$$

$$+ \frac{i}{(l+1)}\left(\frac{(l-l_0+1)(l+l_0+1)(n-l-1)(n+l+1)(l+m+1)(l-m+1)}{(2l+1)(2l+3)}\right)^{1/2} D^{[nl_0]}_{l'ml+1}(\theta),$$

$$[(l \pm m + 1)(l \mp m)]^{1/2}\left(\cos\theta \mp i\frac{l_0n}{l(l+1)}\sin\theta\right)D^{[nl_0]}_{l'ml}(\theta)$$

$$= \frac{i}{l}\left(\frac{(l-l_0)(l+l_0)(n-l)(n+l)(l\pm m+1)(l\pm m)}{(2l-1)(2l+1)}\right)^{1/2}$$

$$\times \sin\theta D^{[nl_0]}_{l'ml-1}(\theta) + [(l' \mp m)(l' \pm m + 1)]^{1/2} D^{[nl_0]}_{l'm\pm 1l}(\theta)$$

$$-\frac{i}{(l+1)}\left(\frac{(l-l_0+1)(l+l_0+1)(n-l-1)(n+l+1)(l\mp m)(l\mp m+1)}{(2l+1)(2l+3)}\right)^{1/2}\sin\theta D^{[nl_0]}_{lm\,l+1}(\theta). \tag{III.6}$$

We next substitute the identity (III.5) into the two Casimir products (III.1) and obtain after some manipulations,

$$\mathbf{L}^2+\mathbf{M}^2=\mathbf{L}^2+M_3^2-\tfrac{1}{2}(F_+F_-+F_-F_+)=M_3^2+2i\cot\theta M_3+\frac{1}{\sin^2\theta}\left[\mathbf{L}^2+\exp(-i\theta M_3)\mathbf{L}^2\exp(i\theta M_3)-2L_3^2\right]$$

$$+L_3^2-\frac{\cot\theta}{\sin\theta}\left[H^\dagger_+\exp(-i\theta M_3)H_+\exp(i\theta M_3)+H^\dagger_-\exp(-i\theta M_3)H_-\exp(i\theta M_3)\right],$$

(III.7)

$$\mathbf{L}\cdot\mathbf{M}=L_3M_3-\frac{i}{2}(H_+F_-+H_-F_+)$$

$$=L_3M_3+i\cot\theta L_3+\frac{i}{2}\frac{1}{\sin\theta}\left[H^\dagger_+\exp(-i\theta M_3)H_+\exp(i\theta M_3)-H^\dagger_-\exp(-i\theta M_3)H_-\exp(i\theta M_3)\right].$$

Multiplying both expressions from the right by $\exp(-i\theta M_3)$ and taking the matrix elements we get after simplifications,

$$\left(\frac{\sin^2\theta}{\cos\theta}\frac{d^2}{d\theta^2}+2\sin\theta\frac{d}{d\theta}+\frac{\sin^2\theta}{\cos^2\theta}a-\frac{1}{\cos\theta}b\right)D^{[nl_0]}_{l'ml}(\theta)$$
$$=-[(l'+m+1)(l'-m)(l+m+1)(l-m)]^{1/2}D^{[nl_0]}_{l'\,m+1\,l}(\theta)$$
$$-[(l'+m)(l'-m+1)(l+m)(l-m+1)]^{1/2}D^{[nl_0]}_{l'\,m-1\,l}(\theta), \tag{III.8}$$

$$\left(m\sin\theta\frac{d}{d\theta}+m\cos\theta+c\sin\theta\right)D^{[nl_0]}_{l'ml}(\theta)$$
$$=-\tfrac{1}{2}[(l'+m+1)(l'-m)(l+m+1)(l-m)]^{1/2}D^{[nl_0]}_{l'\,m+1\,l}(\theta)$$
$$+\tfrac{1}{2}[(l'+m)(l'-m+1)(l+m)(l-m+1)]^{1/2}D^{[nl_0]}_{l'\,m-1\,l}(\theta),$$

$$a=n^2+l_0^2-1-m^2,\quad b=l(l+1)+l'(l'+1)-2m^2,\quad c=il_0n.$$

The five recurrence relations given by (III.6) and (III.8) are sufficient to determine the four raising and lowering ladder operators with respect to $l\,(l')$ and m. In order to obtain similar operators for n and l_0 we have to go outside SO(4). As a simple enlargement, we consider the direct-product group SO(4) ⊗ SO(4) and correspondingly the tensor product of two unitary irreducible SO(4) representations of the principal series T^{nl_0} and $T^{n'l_0'}$, which are realized in the Hilbert spaces $\mathscr{H}^{nl_0}$ and $\mathscr{H}^{n'l_0'}$, spanned respectively by the sets of canonical basis vectors of $\Phi^{nl_0}_{lm}$ and $\Phi^{n'l_0'}_{l_1m_1}$. Naimark[10] has shown that this tensor product representation decomposes as

$$T^{nl_0}\otimes T^{n'l_0'}=\sum_{L_0}\left(\int\sum_N\right)T^{NL_0}\,d\mu(N)$$

[the integration over N with a Haar measure $d\mu(N)$ is for the noncompact extension to SO(3,1) ⊗ SO(3,1)], where the summation over L_0 is such that $L_0+l_0+l_0'=$ nonnegative integer. The representation T^{NL_0} is realized in a Hilbert space $\mathscr{H}^{NL_0}=\mathscr{H}^{nl_0}\oplus\mathscr{H}^{n'l_0'}$ spanned by the set of canonical basis vectors of $\Phi^{NL_0}_{LM}$. From the orthonormality properties of $\Phi^{nl_0}_{lm}\otimes\Phi^{n'l_0'}_{l'm'}$ and $\Phi^{NL_0}_{LM}$ and using (III.2) we, after little manipulation, arrive at the relation,

$$\sum_{l,l'_1}D^{[nl_0]}_{lml'}(\theta)D^{[n'l_0']}_{l'_1m'l'}(\theta)\Phi^{nl_0}_{lm}\otimes\Phi^{n'l_0'}_{l'_1m'}$$
$$=\sum_{NL_0}\sum_{LL'}\sum_{n''l_0''}\sum_{l''m''}\sum_{n'''l_0'''}\sum_{l'''m'''}(\Phi^{nl_0}_{lm}\otimes\Phi^{n'l_0'}_{l'm'},\Phi^{NL_0}_{LM})(\Phi^{n''l_0''}_{l''m''}\otimes\Phi^{n'''l_0'''}_{l'''m'''},\Phi^{NL_0}_{L'M})D^{[NL_0]}_{L'ML}(\theta)\Phi^{n''l_0''}_{l''m''}\otimes\Phi^{n'''l_0'''}_{l'''m'''},$$

$$M=m+m'=m''+m'''. \tag{III.9}$$

We then express the Fano functions $(\Phi^{nl_0}_{lm} \otimes \Phi^{n'l'_0}_{l'm'}, \Phi^{NL_0}_{LM})$ in (III.9) in terms of 3-j and 9-j symbols,[11] multiply both sides of (III.9) by $\begin{pmatrix} l & l' & J \\ m & m' & m+m' \end{pmatrix}$, sum over m and m' and compare the coefficients of the independent basis states. Eventually we obtain (after changing the notation $l_1 \to l'$; $l' \to l_2$ at the end),

$$\exp[i\pi(l-l'+l'_1-l'_2](2L+1)\begin{pmatrix} l & l'_2 & L \\ m & m' & m+m' \end{pmatrix} D^{[n'l'_0]}_{l'm'l'_2}(\theta) D^{[nl_0]}_{lml}(\theta)$$

$$= \sum_{NL_0} \sum_{L'} (N+L_0)(N-L_0)[(2L+1)(2L'+1)(2l+1)(2l'+1)(2l'_1+1)(2l'_2+1)]^{1/2} D^{[NL_0]}_{L'ML}(\theta)$$

$$\times \begin{pmatrix} l' & l'_1 & L' \\ m & m' & m+m' \end{pmatrix} \begin{Bmatrix} \frac{1}{2}(n+l_0-1) & \frac{1}{2}(n'+l_0-1) & \frac{1}{2}(N+L_0-1) \\ \frac{1}{2}(n-l_0-1) & \frac{1}{2}(n'-l_0-1) & \frac{1}{2}(N-L_0-1) \\ l & l'_2 & L \end{Bmatrix}$$

$$\times \begin{Bmatrix} \frac{1}{2}(n+l_0-1) & \frac{1}{2}(n'+l'_0-1) & \frac{1}{2}(N+L_0-1) \\ \frac{1}{2}(n-l_0-1) & \frac{1}{2}(n'-l'_0-1) & \frac{1}{2}(N-L_0-1) \\ l' & l'_1 & L' \end{Bmatrix}. \tag{III.10}$$

For $l'_2 = l'_1 = m' = l'_0 = 0, n' = 2 \Rightarrow L' = l', L = l$ and for $l'_2 = 1, l'_1 = m' = l'_0 = 0, n' = 2 \Rightarrow L' = l', L = l-1, l, l+1$, we get four different relations from (III.10). After using the special values of the Dolginov–Biedenharn function,[6] and the 3-j and 9-j symbols,[12] these relations become:

$$\alpha \cos\theta D^{[nl_0]}_{lml}(\theta) = \alpha_n D^{[n+1\ l_0]}_{lml}(\theta) + \alpha_{-n} D^{[n-1\ l_0]}_{lml}(\theta) + \alpha_{l_0} D^{[n\ l_0+1]}_{lml}(\theta) + \alpha_{-l_0} D^{[n\ l_0-1]}_{lml}(\theta),$$

$$\alpha mi \sin\theta D^{[nl_0]}_{lml}(\theta) = -\alpha_n l_0 D^{[n+1\ l_0]}_{lml}(\theta) + \alpha_{-n} l_0 D^{[n-1\ l_0]}_{lml}(\theta) - \alpha_{l_0} n D^{[n\ l_0+1]}_{lml}(\theta) + \alpha_{-l_0} n D^{[n\ l_0-1]}_{lml}(\theta),$$

$$\beta i \sin\theta D^{[nl_0]}_{lm\ l-1}(\theta) = \beta^{l_0}_n D^{[n+1\ l_0]}_{lml}(\theta) - \beta^{l_0}_{-n} D^{[n-1\ l_0]}_{lml}(\theta) - \beta^{n}_{l_0} D^{[n\ l_0+1]}_{lml}(\theta) - \beta^{n}_{-l_0} D^{[n\ l_0-1]}_{lml}(\theta),$$

$$\gamma i \sin\theta D^{[nl_0]}_{lm\ l+1}(\theta) = -\gamma^{l_0}_n D^{[n+1\ l_0]}_{lml}(\theta) + \gamma^{l_0}_{-n} D^{[n-1\ l_0]}_{lml}(\theta) - \gamma^{n}_{l_0} D^{[n\ l_0+1]}_{lml}(\theta) - \gamma^{n}_{-l_0} D^{[n\ l_0-1]}_{lml}(\theta),$$

$$\alpha = 2(n-l_0)(n+l_0),$$

$$\beta = \alpha[(2l+1)(l+m)(l-m)]^{1/2},$$

$$\gamma = -\alpha[(2l+1)(l+m+1)(l-m+1)]^{1/2},$$

$$\alpha_x = [(x+l+1)(x-l)(x+l'+1)(x-l')]^{1/2},$$

$$\beta^y_x = [(2l-1)(x+l)(x+l+1)(x+l'+1)(x-l')(l+y)(l-y)]^{1/2},$$

$$\gamma^y_x = [(2l+3)(x-l-1)(x-l)(x+l'+1)(x-l')(l+y+1)(l-y+1)]^{1/2}. \tag{III.11}$$

The relations (III.6), (III.8), and (III.11) exactly agree with those obtained by Ström[8] who made use of an explicit realization of SO(4) generators in terms of differential operators in the six group parameters. From these nine recurrence relations (III.6), (III.8), and (III.11) we obtain the following eight ladder operators:

$$L^{\pm}_m D^{[nl_0]}_{lml}(\theta) = \Delta^{0}_{l\pm m} D^{[nl_0]}_{lml}(\theta) = D^{[nl_0]}_{l'm\pm 1\ l}(\theta), \quad L^{+}_{l} D^{[nl_0]}_{lml}(\theta) = \Delta^{-(l+1)}_{+l} D^{[nl_0]}_{lml}(\theta) = D^{[nl_0]}_{l'm\ l\pm 1}(\theta), \quad L^{+}_{l'} = L^{+}_{l-l'},$$

$$L^{\pm}_n D^{[nl_0]}_{lml}(\theta) = \Delta^{0}_{l\pm n} D^{[nl_0]}_{lml}(\theta) = D^{[n\pm 1\ l_0]}_{lml}(\theta), \quad L^{\pm}_{l_0} D^{[nl_0]}_{lml}(\theta) = \Delta^{0}_{l\pm l_0} D^{[nl_0]}_{lml}(\theta) = D^{[nl_0\pm 1]}_{lml}(\theta), \tag{III.12}$$

$$\lambda^{\mu}_x \sin\theta \Delta^{\mu}_x = \frac{1}{2}\frac{\sin^2\theta}{\cos\theta}\frac{d^2}{d\theta^2} + (x+1)\sin\theta\frac{d}{d\theta} + \frac{a}{2}\frac{\sin^2\theta}{\cos\theta} - \frac{b}{2\cos\theta} + cm\sin\theta\frac{1}{x} + [x(x+1)-m^2]\cos\theta,$$

$$\lambda^{\mu}_x = \left(\frac{i}{x}\right)^{\mu}\left(\frac{2x+1}{2x-1}(m+x)(x-m)(l_0+x)(x-l_0)(n+x)(n-x)\right)^{\mu/2}[(l+x+1)(l-x)(l'+x+1)(l'-x)]^{(1-\mu)/2}.$$

These ladder operators satisfy the operator equations

$$[L^{\mp}_{r\pm 1} L^{+}_{r} - \mathbb{1}] D^{[nl_0]}_{lml}(\theta) = 0, \quad r = (m,l,n,l_0), \tag{III.13}$$

which in turn are equivalent to the fourth order differential equation,

$$\left[a_4(\theta)\left(\sin^4\theta\frac{d^4}{d\theta^4}\right) + a_3(\theta)\left(\sin^3\theta\frac{d^3}{d\theta^3}\right) + a_2(\theta)\left(\sin^2\theta\frac{d^2}{d\theta^2}\right) + a_1(\theta)\left(\sin\theta\frac{d}{d\theta}\right) + a_0(\theta)\right] D^{[nl_0]}_{lml}(\theta) = 0,$$

$$a_4(\theta) = \frac{1}{\cos^2\theta}, \quad a_3(\theta) = \frac{2}{\cos^3\theta} + \frac{4}{\cos\theta}, \quad a_2(\theta) = 2\tan^4\theta + (4+2a-2b)\tan^2\theta + (6-2m^2-2b),$$

$$a_1(\theta) = 2a\frac{\sin^4\theta}{\cos^3\theta} - 2b\frac{\sin^2\theta}{\cos^3\theta} - 8mc\sin\theta + \frac{1}{\cos\theta}(4m^2 + 6a - 2b) - \cos\theta(8m^2 + 6a),$$

$$a_0(\theta) = (2a - 2b)\tan^4\theta + (4m^2 - a - 4c^2)\sin^2\theta - 8mc\sin\theta\cos\theta$$

$$+ 4mc\tan\theta + [(a - b + 1)^2 - 1]\frac{1}{\cos^2\theta} + [-(a - b + 1)^2 + (b + 2m^2)^2 + 1 - 4ll'(l + 1)(l' + 1)]. \tag{III.14}$$

The solution of this equation, finite for $\cos\theta = \pm 1$ is given by (III.3). Also, the four different ways of factorization of this equation in terms of second-order differential equations are given by the operator equations (III.13). We find that out of the four relations (III.12) only the ladder operators $L_l^{\pm}$ corresponding to the angular momentum l differ from the others mainly by an extra θ dependence, since $\mu = 1$ introduces an extra $\sin\theta$ for $L_l^{\pm}$. This situation is similar to the case considered earlier in Sec. II.

The ladder operators for SO(3,1) can be obtained in the similar fashion. The six generators L_i, M_i of SO(3,1) satisfy the Lie-commutation

$$[L_i,L_j] = i\epsilon_{ijk}L_k, \quad [M_i,M_j] = -i\epsilon_{ijk}M_k, \quad [L_i,M_j] = i\epsilon_{ijk}M_k.$$

In a unitary irreducible representation of SO(3,1) belonging to the principal series one constructs the orthonormal basis states $\Phi_{lm}^{\nu l_0}$, $-\infty < \nu < +\infty$, such that the eigenvalues of the invariant Casimir products are

$$(\mathbf{L}^2 - \mathbf{M}^2)\Phi_{lm}^{\nu l_0} = (-\nu^2 + l_0^2 - 1)\Phi_{lm}^{\nu l_0}, \quad (\mathbf{L}\cdot\mathbf{M})\Phi_{lm}^{\nu l_0} = \nu l_0\Phi_{lm}^{\nu l_0}. \tag{III.15}$$

The Dolginov–Toptygin function[13] is defined by the matrix elements (the matrix elements of a finite Lorentz boost transformation in the third direction)

$$(\Phi_{l'm}^{\nu l_0}, \exp(-i\theta M_3)\Phi_{lm}^{\nu l_0}) \equiv D_{l'ml}^{[i\nu l_0]}(\theta). \tag{III.16}$$

An integral form of this matrix element is obtained as

$$D_{l'ml}^{[i\nu l_0]}(\theta) = \tfrac{1}{2}[(2l' + 1)(2l + 1)]^{1/2}\left(\frac{(i\nu + l')!(i\nu - l' - 1)!}{(i\nu + l)!(i\nu - l - 1)!}\right)^{1/2}\int_{-1}^{+1} dx\,(\cosh\theta - x\sinh\theta)^{i\nu - 1}$$

$$\times d_{l_0m}^{l'*}(x)d_{l_0m}^{l}\left(\frac{x\cosh\theta - \sinh\theta}{\cosh\theta - x\sinh\theta}\right), \tag{III.17}$$

$$\left(\frac{(i\nu + l')!(i\nu - l' - 1)!}{(i\nu + l)!(i\nu - l - 1)!}\right)^{1/2} = \prod_{k=l_0}^{k=l}\frac{(-k + i\nu)}{(k^2 + \nu^2)}\prod_{k'=l_0}^{k'=l'}\frac{(-k' - i\nu)}{(k'^2 + \nu^2)}.$$

This integral form can be converted into a series as in (III.4) with $l_{\pm} = \frac{1}{2}(i\nu - 1) \pm \frac{1}{2}l_0$.

All recurrence relations and ladder operators of the Dolginov–Toptygin function can be trivially obtained from those for SO(4) by means of the analytical continuation $\theta \to -i\theta$ and $n \to i\nu$. This is also transparent if we compare (III.3) with (III.17). We give below all the eight operators:

$$L_m^{\pm}D_{l'ml}^{[i\nu l_0]}(\theta) = \Delta_{[\pm m]}^{0}D_{l'ml}^{[i\nu l_0]}(\theta) = D_{l'm\pm 1\,l}^{[i\nu l_0]}(\theta), \quad L_l^{\pm}D_{l'ml}^{[i\nu l_0]}(\theta) = \Delta_{[\substack{-(l+1)\\+l}]}^{\prime}D_{l'ml}^{[i\nu l_0]}(\theta) = D_{l'm\,l\pm 1}^{[i\nu l_0]}(\theta),$$

$$L_\nu^{\pm}D_{l'ml}^{[i\nu l_0]}(\theta) = \Delta_{[\pm i\nu]}^{0}D_{l'ml}^{[i\nu l_0]}(\theta) = D_{l'ml}^{[i\nu\pm 1\ l_0]}(\theta), \quad L_{l_0}^{\pm}D_{l'ml}^{[i\nu l_0]}(\theta) = \Delta_{[\pm l_0]}^{0}D_{l'ml}^{[i\nu l_0]}(\theta) = D_{l'ml}^{[i\nu l_0\pm 1]}(\theta), \tag{III.18}$$

$$\lambda_x^{\mu}(-i\sinh\theta)^{\mu}\Delta_x^{\mu} = \frac{1}{2}\frac{\sinh^2\theta}{\cosh\theta}\frac{d^2}{d\theta^2} + (x + 1)\sinh\theta\frac{d}{d}$$

$$- \frac{\sinh^2\theta}{2\cosh\theta}a' - \frac{b'}{2\cosh\theta} + ic'm\,\sinh\theta\frac{1}{x} + [x(x + 1) - m^2]\cosh\theta,$$

$$a' = -\nu^2 + l_0^2 - 1 - m^2, \quad b' = l(l + 1) + l'(l' + 1) - 2m^2, \quad c' = l_0\nu.$$

These operators, as in previous cases, satisfy the operator equations (III.13) with $r = (m,l,i\nu,l_0)$ which can be transformed into a fourth-order differential equation as in (III.14) with $\theta \to -i\theta$, $a \to a'$, $b \to b'$, and $c \to c'$. The four types of factorization are also clear in this case.

Finally, we deal with the Euclidean group E(3). We define the six generators L_i, M_i such that they satisfy the commutation relations

$$[L_i,L_j]=i\epsilon_{ijk}L_k,\quad [M_i,M_j]=0,\quad [L_i,M_j]=i\epsilon_{ijk}M_k.$$

In other words, L_i and M_i are respectively the generators of the SO(3) and T_3—the Abelian three-translation—, subgroups of E(3). In a unitary irreducible representation of E(3) one can construct the discrete (spherical) orthonormal basis states $\Phi^{l_0}_{lm}(\tau)$, $0<\tau^2<\infty$, $l_0=0,\pm\frac{1}{2},\cdots$, $l=|l_0|,1+|l_0|,\ldots,\infty$, $-l\leqslant m\leqslant +l$ so that the eigenvalues of the two invariant Casimir products are:

$$\mathbf{M}^2\Phi^{l_0}_{lm}(\tau)=\tau^2\Phi^{l_0}_{lm}(\tau),\quad (\mathbf{L}\cdot\mathbf{M})\Phi^{l_0}_{lm}(\tau)=\tau l_0\Phi^{l_0}_{lm}(\tau), \tag{III.19}$$

and (l,m) are the eigenvalues belonging to the SO(3) subgroup.

The Vilenkin–Akim–Levin function[14] in E(3) is defined by the matrix element (M_3 is the translation operator in the z direction),

$$(\Phi^{l_0}_{l'm}(\tau),\exp(-izM_3)\Phi^{l_0}_{lm}(\tau))\equiv J^{l_0}_{l'ml}. \tag{III.20}$$

This function, as all other functions defined earlier, has a simple integral representation,

$$J^{l_0}_{l'ml}(\tau z)=\tfrac{1}{2}[(2l'+1)(2l+1)]^{1/2}\int_{-1}^{+1}dx\,\exp[-i\tau zx]d^{l'*}_{l_0m}(x)d^{l}_{l_0m}(x), \tag{III.21}$$

which can also be converted into a series form,

$$J^{l_0}_{l'ml}(\tau z)=\tfrac{1}{2}[(2l'+1)(2l+1)]^{1/2}\sum_L i^L(2l'+1)(2L+1)\begin{pmatrix}l' & l & L\\ m & -m & 0\end{pmatrix}\begin{pmatrix}l' & l & L\\ l_0 & -l_0 & 0\end{pmatrix}j_L(\tau z), \tag{III.22}$$

where $j_L(\tau z)$ is the spherical Bessel functions.

In order to compute the ladder operators of $J^{l_0}_{l'ml}(\tau z)$ one can follow the method we have adopted for SO(4). However, the same results can be trivially obtained from those for SO(4) by means of a limiting procedure (Inönü–Wigner contraction) as in the case of E(2) where the results can be obtained as a limit of those for SO(3).

We define six generators $L'_i=L_i$, $M'_i=M_i\sqrt{\epsilon}$ belonging to a "master group" G(4;ϵ) in such a way that L_i, M_i from the generators of SO(4). The primed generators satisfy the commutations:

$$[L'_i,L'_j]=i\epsilon_{ijk}L'_k,\quad [M'_i,M'_j]=\epsilon i\epsilon_{ijk}M'_k,\quad [L'_i,M'_j]=i\epsilon_{ijk}M'_k.$$

for $\epsilon=+1,-1,0$ we respectively restore the algebra of SO(4), SO(3,1), and E(3). The two Casimir products of G(4;ϵ) are

$$\mathbf{L}^2+\epsilon\mathbf{M}^2=(x^2+l_0^2-1)\mathbb{1},\quad \sqrt{\epsilon}\mathbf{L}\cdot\mathbf{M}=xl_0\mathbb{1},$$

so that for $\epsilon=+1\Rightarrow x=n$, $\epsilon=-1\Rightarrow x=iv$ and in the limit $\sqrt{\epsilon}\to 0$ and $|x|\to\infty$ with $x\sqrt{\epsilon}=\tau$, they go over to (III.1), (III.15), and (III.19), respectively. Furthermore, under this limiting procedure one[14] can easily obtain (III.21) and (III.22) from (III.3) and (III.4) in virtue of the Brussard–Tolhock asymptotic formula[6] for the 3-j symbol and the Bessel expansion formula

$$\exp(iy\cos\theta)=\sum_n i^n(2n+1)j_n(y)d^n_{00}(\theta).$$

The recurrence relations and the ladder operators of $J^{l_0}_{l'ml}(\tau z)$ can be readily obtained from those of the Dolginov–Biedenharn functions by using the above limiting procedure and taking $\theta=z\sqrt{\epsilon}$. We find,

$$\left(\frac{1}{\tau}\frac{d}{dz}+i\frac{ml_0}{l(l+1)}\right)J^{l_0}_{l'ml}(\tau z)$$
$$=\frac{i}{(l+1)}\left(\frac{(l-l_0+1)(l+l_0+1)(l+m+1)(l-m+1)}{(2l+1)(2l+3)}\right)^{1/2}J^{l_0}_{l'm\,l+1}(\tau z)$$
$$+\frac{i}{l}\left(\frac{(l-l_0)(l+l_0)(l-m)(l+m)}{(2l-1)(2l+1)}\right)^{1/2}J^{l_0}_{l'm\,l-1}(\tau z),$$

$$[(l\pm m+1)(l\mp m)]^{1/2}\left(1\mp i\frac{l_0\tau z}{l(l+1)}\right)J^{l_0}_{l'ml}(\tau z)$$
$$=\frac{i\tau z}{l}\left(\frac{(l-l_0)(l+l_0)(l\pm m+1)(l\pm m)}{(2l-1)(2l+1)}\right)^{1/2}J^{l_0}_{l'm\,l-1}(\tau z)+[(l'\mp m)(l'\pm m+1)]^{1/2}J^{l_0}_{l'm\pm1\,l}(\tau z)$$
$$-\frac{i\tau z}{(l+1)}\left(\frac{(l-l_0+1)(l+l_0+1)(l\mp m)(l\mp m+1)}{(2l+1)(2l+3)}\right)^{1/2}J^{l_0}_{l'm\,l+1}(\tau z),$$

$$\left(z^2\frac{d^2}{dz^2}+2z\frac{d}{dz}+\tau^2z^2+2m^2-l(l+1)-l'(l'+1)\right)J^{l_0}_{l'ml}(\tau z)$$
$$=-[(l'+m+1)(l'-m)(l+m+1)(l-m)]^{1/2}J^{l_0}_{l'\ m+1\ l}(\tau z)$$
$$-[(l'+m)(l'-m+1)(l+m)(l-m+1)]^{1/2}J^{l_0}_{l'\ m-1\ l}(\tau z),$$

$$\left(mz\frac{d}{dz}+m+il_0\tau z\right)J^{l_0}_{l'ml}(\tau z)$$
$$=-\tfrac{1}{2}[(l'+m+1)(l'-m)(l+m+1)(l-m)]^{1/2}J^{l_0}_{l'\ m+1\ l}(\tau z)$$
$$+\tfrac{1}{2}[(l'+m)(l'-m+1)(l+m)(l-m+1)]^{1/2}J^{l_0}_{l'\ m-1\ l}(\tau z). \tag{III.23}$$

Similarly the six ladder operators are:

$$L^{\pm}_{m}J^{l_0}_{l'ml}(\tau z)=\Delta^{0}_{[\pm m]}J^{l_0}_{l'ml}(\tau z)=J^{l_0}_{l'\ m\pm1\ l}(\tau z),\quad L^{\pm}_{l}J^{l_0}_{l'ml}(\tau z)=\Delta^{1}_{\left[\begin{smallmatrix}-(l+1)\\+l\end{smallmatrix}\right]}J^{l_0}_{l'ml}(\tau z)=J^{l_0}_{l'm\ l\pm1}(\tau z),$$

$$L^{\pm}_{l_0}J^{l_0}_{l'ml}(\tau z)=\Delta^{0}_{[\pm l_0]}J^{l_0}_{l'ml}(\tau z)=J^{l_0\pm1}_{l'ml}(\tau z),$$

$$\lambda^{\mu}_{x}(\tau z)^{\mu}\Delta^{\mu}_{x}=\tfrac{1}{2}z^2\frac{d^2}{dz^2}+(x+1)z\frac{d}{dz}+\tfrac{1}{2}\tau^2z^2+i\frac{1}{x}ml_0\tau z-\tfrac{1}{2}[l(l+1)+l'(l'+1)-2m^2]+[x(x+1)-m^2], \tag{III.24}$$

$$\lambda^{\mu}_{x}=[(l+x+1)(l-x)(l'+x+1)(l'-x)]^{(1-\mu)/2}\left(\frac{i}{x}\right)^{\mu}\left(\frac{2x+1}{2x-1}(m^2-x^2)(l_0^2-x^2)\right)^{\mu/2}.$$

The operators $L^{\pm}_{m}$ and $L^{\pm}_{l}$ are also obtained by Ström.[14] Again, these ladder operators satisfy the operator equations,

$$[L^{\mp}_{r+1}L^{\pm}_{r}-\mathbb{1}]J^{l_0}_{l'ml}(\tau z)=0,\quad r=(m,l,l_0), \tag{III.25}$$

which are equivalent to the fourth-order differential equation,

$$\left[\alpha_4\left(z^4\frac{d^4}{dz^4}\right)+\alpha_3\left(z^3\frac{d^3}{dz^3}\right)+\alpha_2(\tau z)\left(z^2\frac{d^2}{dz^2}\right)+\alpha_1(\tau z)\left(z\frac{d}{dz}\right)+\alpha_0(\tau z)\right]J^{l_0}_{l'ml}(\tau z)=0,$$
$$\alpha_4=1,\quad \alpha_3=6,\quad \alpha_2(\tau z)=2\tau^2z^2+2m^2-2l(l+1)-2l'(l'+1)+6, \tag{III.26}$$
$$\alpha_1(\tau z)=-8iml_0\tau z-2l(l+1)-2l'(l'+1),\quad \alpha_0(\tau z)=(4l_0^2-1)\tau^2z^2-4iml_0\tau z+[l(l+1)-l'(l'+1)]^2.$$

The operator equations (III.25) indicate the three different ways of factorizing Eq. (III.26). As in the previous cases $L^{\pm}_{l}$ differ from the others through an extra overall multiplication by z.

In conclusion, we have constructed all ladder operators of the transition matrix elements defined in the representation spaces of SO(3), SO(2,1), E(2); SO(4), SO(3,1), and E(3) by considering also the direct product space of SO(3) ⊗ SO(3), etc. We have noticed that the ladder operators corresponding to the angular momentum l (or K) of the (sub) group SO(3) [or SO(2,1)] differ mainly from the others through an overall multiplication factor that depends on the group parameter θ. Furthermore, using these ladder operators one can in principle determine all recurrence relations satisfied by the respective matrix elements. Also, we have seen how a typical second- and fourth-order differential equation can be factorized, as an extension to the Schrödinger–Infeld–Miller factorization, in terms of first- and second-order ladder operators and how additional factorizations can be achieved by considering the direct product groups. Finally, as pointed out earlier these ladder operators play crucial roles in physical calculations where these matrix elements describe physical transitions.[6]

ACKNOWLEDGMENTS

We wish to thank Professor A.O. Barut for many illuminating discussions and for his strong interest in this work. One of us (R.W.) is thankful to Professor L. Castell and the Physik Gruppe at Starnberg for their kind hospitality and some financial support. We are very thankful to Miss Traudi Lehmeier for typing this manuscript. We are very grateful to the referee for kindly pointing out two errors and their corrections.

[1]E. Schrödinger, Proc. Roy. Irish Acad. A **46**, 9 (1940); **46**, 183 (1941); **47**, 53 (1941); L. Infeld and T.E. Hull, Rev. Mod. Phys. **23**, 21 (1951); T. Inui, Prog. Theor. Phys. **3**, 168, 244 (1948).
[2]W. Miller, Jr., Mem. Amer. Math. Soc. **50**, 5 (1964); B. Kaufman, J. Math. Phys. **7**, 447 (1966); Masao Mori, J. Math. Phys., **19**, 1057 (1978).

[3]Y. Miyachi and Y. Takao, Math. Jpn. **22**, 369 (1977).
[4]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*, American Math. Soc. Monograph Vol. 22 (Am. Math. Soc., Providence, R.I., 1968).
[5]V. Bargmann, Ann. Math. **48**, 568 (1974); G. Lindblad and B. Nagel, Ann. Inst. Henri Poincaré A **13**, 27 (1970).
[6]A.O. Barut and Raj Wilson, J. Math. Phys. **17**, 900 (1976).
[7]E. Inönü and E.P. Wigner, Proc. Nat. Acad. Sci. USA. **39**, 510 (1953); W.T. Sharp, Research Report, CRT-935, AECL-1098, Atomic Energy of Canada, Chalk River, Ontario, 1960.
[8]A.Z. Dolginov, Ż. Èksp. Teor. Fiz. **30**, 746 (1956) [Sov. Phys. JETP **3**, 589 (1956)]; L.C. Biedenharn, J. Math. Phys. **2**, 433 (1961); S. Ström, Ark. Fysik **40**, 1 (1968).
[9]M.A. Naimark, Usp. Mat. Nauk **9**, 19 (1954) [Amer. Math. Soc. Transl. (2) **6**, 379 (1957)].
[10]M.A. Naimark, Trudy Moskov Mat. Obšč. **8**, 121 (1959); **9**, 237 (1960); **10**, 181 (1961) [Amer. Math. Soc. Transl. (2) **36**, 101, 137, 189 (1964)]; R.L. Anderson, R. Rączka, M.A. Rashid, and P. Winternitz, J. Math. Phys. **11**, 1050, 1059 (1970); F.L. Williams, Lecture Notes in Math. No. 358 (Springer-Verlag, Berlin, 1973).
[11]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[12]H. Matsunobu and H. Takebe, Prog. Theor. Phys. **14**, 589 (1955).
[13]A.Z. Dolginov and I.N. Toptygin, Ž. Èksp. Teor. Fiz. **35**, 794 (1958) [Sov. Phys. JETP **8**, 550 (1959)]; W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).
[14]N. Ja. Vilenkin, E.L. Akim, and A.A. Levin, Dokl. Akad. Nauk SSSR **112**, 987 (1957) (in Russian); W. Miller, Jr., Commun. Pure App. Math. **17**, 527 (1964); S. Ström, Ark. Fysik **30**, 267 (1965); W.J. Holman III, Ann. Phys. (N.Y.) **52**, 176 (1969).

On the multiplicity-free Wigner and Racah coefficients of U(n)

M. K. F. Wong
Fairfield University, Fairfield, Connecticut 06430

(Received 2 November 1978; accepted for publication 22 June 1979)

A necessary and sufficient condition is found for the tensor operators of U(n) to be multiplicity-free. At present all matrix elements for the multiplicity-free tensor operators of U(n) are known explicitly in closed form. The phase convention given by Wong in a previous paper is modified slightly. Some inconsistencies in the equations found in the two previous papers of Wong regarding phase factors are corrected. The present phase convention for the multiplicity-free Wigner and Racah coefficients of U(n) is the most general extension of the Condon–Shortley phase convention from U(2) to U(n).

I. INTRODUCTION

At present it seems that a canonical solution to the multiplicity problem of the Wigner coefficients of U(n) for $n > 3$ is rather difficult, especially after Derome[1,2] and subsequent workers[3] have proved that a simple phase relation for the general Wigner coefficients of U(n), $n \geqslant 4$, under permutation of any two columns is impossible. We must therefore be content to consider the simpler problem first, viz., the multiplicity-free Wigner coefficients of U(n). In this connection there are at least two questions which should be settled. First, what are the necessary and sufficient conditions for a tensor operator of U(n) to be multiplicity free? Second, can the matrix elements of the multiplicity-free tensor operators of U(n) be explicitly written down in closed form? Since, by the Wigner–Eckart theorem, the matrix elements of the tensor operators are closely connected with the Wigner coefficients, a practical problem in connection with the second question is the phase convention to be adopted for the multiplicity-free Wigner coefficients of U(n).

In 1976, Wong[4,5] (henceforth referred to as I and II) gave a phase convention for the multiplicity-free Wigner coefficients of U(n), which is an extension of the phase convention from U(2) to U(n). This phase convention has the advantage that one can immediately check the results for U(n) by first applying it to U(2) and seeing whether it agrees with the conventional Condon–Shortley phase in the simplest case. The second advantage of this phase convention is that under conjugation, the phases behave in the same way as U(2), i.e., $j^* = j$, $m^* = -m$.

Basically, this phase convention makes the following mapping:

$$j \to \epsilon_n (m_{1n} - m_{nn}),$$

$$m \to \tfrac{1}{2} \sum_{i=1}^{n-1} z_i,$$

where ϵ_n and z_i are suitably defined quantities.

There are, however, several other conditions which the previous phase convention did not take into consideration. One of these is that the phase factor should always be an integer. Basically, the mapping above for j is too simple. Though it works for U(2) and U(3), it does not work for U(n), $n \geqslant 4$. Thus in this paper we wish to modify the mapping of j slightly, for $n \geqslant 4$.

In Sec. III we shall give the revised phase convention, which, as we shall show, will satisfy seven requirements. These requirements, as far as we can ascertain at this moment, encompass practically all the conditions one can impose on the phase factor.

In I and II, Wong has shown that the multiplicity-free 3-j, 6-j, and 9-j symbols of U(n) and U($n-1$) are all related to each other. As a result he obtained in II an explicit evaluation of the multiplicity-free 6-j symbol of U(n). Unfortunately, there have been some typographical errors with regard to the dimensional factors, and some inconsistencies in the phase factors of the equations in I and II. We take this opportunity to correct these mistakes. This results in even simpler phase factors in the equations involved. We hope that this revised phase convention for U(n) will be adopted as the standard extension of the Condon–Shortley phase convention, so that practical calculations can be made on the basis of this convention.

In Sec. II we settle a question raised in II, i.e., what is the necessary and sufficient condition for a tensor operator of U(n) to be multiplicity free. After answering the question, we then conclude that the matrix elements of all multiplicity-free tensor operators of U(n) are now known in closed form. In Sec. III we give the phase factors and symmetry properties for the multiplicity-free Wigner coefficients of U(n). In Sec. IV we give the phase factors and symmetry properties for the multiplicity-free Racah coefficients of U(n).

II. NECESSARY AND SUFFICIENT CONDITIONS FOR THE TENSOR OPERATORS OF U(N) TO BE MULTIPLICITY FREE

We start by asking the following question: Take the direct product of

$$[m_1] \otimes [m_2] \to \tilde{y}(m_3)[m_3]$$

in U(n), where $[m_1]$ is an arbitrary representation, and $\tilde{y}(m_3)$ is the outer multiplicity of the irreducible representation $[m_3]$. What form must $[m_2]$ take so that $\tilde{y}(m_3)$ is equal to one? We answer the question as follows: $[m_2]$ is either of the form $[p,\dot{0}]$ (or its conjugate $[\dot{p},0]$), or of the form [1...1 0...0]. In the following we shall also show that this is a necessary and sufficient condition for a tensor operator of U(n) to be multiplicity free.

 0022-2488/79/122391-07$01.00

First we prove that this is a sufficient condition, i.e., if $[m_2]$ is of the form $[p,\dot{0}]$, $[\dot{p},0]$, or [1...1 0...0], then $\tilde{y}(m_3) = 1$ for $[m_1]$ arbitrary. One simply forms the direct product of $[m_1]$ with any of the three representations mentioned above and finds that the resulting $[m_3]$ is multiplicity free, i.e., $\tilde{y}(m_3) = 1$.

Another way to prove the sufficiency condition is to use the Racah–Speiser lemma:

$$\tilde{y}(m_3) = \sum_S \sum_m y(m)\delta_S \delta_{S(m_1 + m + R_0), m_3 + R_0}, \tag{2.1}$$

where $\tilde{y}(m_3)$ is the outer multiplicity for the irreducible representation $[m_3]$; $y(m)$ is the inner multiplicity of m belonging to the irreducible representation $[m_2]$; S = operation of the Weyl group; $\delta_S = +1$ or -1 according to whether S is even or odd, respectively; R_0 = half the sum of all the positive roots of U(n); and the second δ is a Kronecker delta.

From (2.1) one sees that the outer multiplicity of $[m_3]$ in our problem, $\tilde{y}(m_3)$, cannot be greater than $y(M)$, where $y(M)$ is the highest inner multiplicity of a certain weight M belonging to $[m_2]$. The proof of this statement is given in the Appendix. Thus the problem of finding the sufficient condition is equivalent to the following problem: Find all the irreducible representation in U(n) where all the inner multiplicities of weights are equal to one. We find that for the irreducible representations in U(n): $[p,\dot{0}]$, $[\dot{p},0]$, and [1...1 0...0], all have inner multiplicity of weights equal to one. Moreover, these are the only irreducible representations having this property. To show this, one first checks that for U(3), the next irreducible representation to be considered is [2,1,0], which is the well-known octet representation, and has inner multiplicity 2 for the weight [0,0,0]. For U(4), one finds that the irreducible representations [2,1,0,0], [2,1,1,0], and [2,2,0,0] all have at least one weight whose inner multiplicity is greater than 1. The other representations in U(n), $n \geqslant 4$, will have at least equal or more complicated inner multiplicity structure than those in U(3) and U(4). Hence we conclude that these three representations, $[p,\dot{0}]$, $[\dot{p},0]$, [1...1 0...0], are the only ones whose inner multiplicity of weights are all equal to one. It turns out that they are not only the sufficient, but also the necessary conditions, for a tensor operator of U(n) to be multiplicity free.

Now by necessity, we mean the following: Given $[m_2]$, not one of the three mentioned above, then there exists at least one $[m_1]$, such that for $[m_1] \otimes [m_2] \rightarrow \tilde{y}(m_3)[m_3]$ there is at least one $[m_3]$ where $\tilde{y}(m_3) > 1$.

For the necessary condition, here is an outline of the proof by a process of enumeration. For U(3), given $[m_2]$ satisfying the conditions above, we can choose $[m_1]$ to be the adjoint representation [2,1,0]. Then the resulting $[m_3] = [m_2]$ has $\tilde{y}(m_3) = 2$.

For U(4), again let us choose $[m_1]$ to be the adjoint representation [2,1,1,0]. Then we find that the only kind of $[m_2]$ other than the three above which will be multiplicity free is of the form $[p,p,0,0]$. Hence if $[m_2]$ is not of this form, then we choose $[m_1]$ to be [2,1,1,0] and find for $[m_3] = [m_2]$, $\tilde{y}(m_3) > 1$. Now for $[m_2] = [p,p,0,0]$, we can choose $[m_1]$ to be, say, [3,2,1,0]. Then for $[m_3] = [p+2, p+1, 2, 1]$, we find $\tilde{y}(m_3) > 1$.

This is also true for U(n), i.e., for $[m_2] = [p,p,\dot{0}]$, $[m_1] = [3,2,1,\dot{0}]$, and $[m_3] = [p+2, p+1, 2, 1, \dot{0}]$, $\tilde{y}(m_3) > 1$ for U(n).

Now for U(n) in general, we take $[m_1]$ to be the adjoint representation [2,1,...,1,0]. Then we find that the only irreducible representations other than the three mentioned above which are multiplicity free are of the form [p,..., p, 0,...,0]. Let us say that there are i p's in $[m_2]$, where, without loss of generality, $2i \leqslant n$, since otherwise we can always consider its conjugate, which has the same structure. Then we choose $[m_1]$ to be $[2i-1, 2i-2, ..., 1, \dot{0}]$. We then find that for $[m_3] = [p+2i-2, p+2i-3, ..., p+i-1, i, i-1, ..., 1, \dot{0}]$, $\tilde{y}(m_3) > 1$.

Thus we have proved the following statement: Given $[m_2] = [p,\dot{0}]$, $[\dot{p},0]$, or [1...1 0...0], then $\tilde{y}(m_3) = 1$ for all $[m_1]$ and $[m_3]$. Given $[m_2]$ not equal to one of the three representations mentioned above, then there is at least one $[m_1]$ such that in the direct product $[m_1] \otimes [m_2] \rightarrow \tilde{y}(m_3)[m_3]$, there exists at least one $[m_3]$ where $\tilde{y}(m_3) > 1$.

It is also clear from what we have said above that the necessary condition does not mean the following: Given $\tilde{y}(m_3) = 1$, then $[m_2]$ is one of the three forms mentioned above. This statement is obviously not true since for

$$[m_{1n}, m_{2n}, ..., m_{nn}] \otimes [m'_{1n}, m'_{2n}, ..., m'_{nn}] \rightarrow [m_{1n} + m'_{1n}, m_{2n} + m'_{2n}, ..., m_{nn} + m'_{nn}],$$

$\tilde{y}(m_3)$ is always equal to one.

It is also interesting to note that there are cases where both $[m_1]$ and $[m_2]$ are not multiplicity free as far as the inner multiplicities are concerned, but their direct product is multiplicity free. For example, in U(4),

$$[2,1,1,0] \otimes [2,2,0,0] = [4,3,1,0] + [3,3,2,0] + [2,2,0,0] + [2,1,1,0] + [3,1,0,0],$$

where both [2,1,1,0] and [2,2,0,0] are not multiplicity free as far as the inner multiplicity of weights is concerned.

We thus conclude that the three representations mentioned above are the only ones that give rise to multiplicity-free Wigner coefficients all the time, and in this sense they merit being called the necessary and sufficient conditions for a tensor operator of U(n) to be multiplicity free.

The Wigner coefficients U(n) involving the totally symmetric representation $[p,\dot{0}]$ or its conjugate $[\dot{p},0]$ are well-known. (See, e.g., references quoted in I.) The irreducible representation $[1_k, \dot{0}_{n-k}]$ are called by Biedenharn and Louck[6] the "elementary representation." The corresponding tensor operators are called the elementary operators. Their matrix elements have been explicitly obtained in Ref. 6. We can therefore say that the matrix elements for all multiplicity-free tensor operators of U(n) have been explicitly obtained in closed form.

Let us give the results for the matrix elements of the elementary operators explicitly,

$$\begin{bmatrix} (1,2\cdots k) \\ [\dot{1}_k \cdots \dot{0}_{n-k}] \\ (1,2\cdots k) \end{bmatrix} = + \left[\prod_{i=1}^{k} \left(\prod_{j=k+1}^{n} \frac{(p_{i,n-1} - p_{jn} + 1)}{(p_{in} - p_{jn})} \prod_{j=k+1}^{n-1} \frac{(p_{in} - p_{j,n-1})}{(p_{i,n-1} - p_{j,n-1} + 1)} \right) \right]^{1/2}, \tag{2.2}$$

$$\left\langle \begin{pmatrix} [m]_n + \Delta(i_1,i_2,\ldots,i_k) \\ [m]_{n-1} + \Delta(j_1,j_2,\ldots,j_k) \end{pmatrix} \middle\| \begin{bmatrix} (i_1 & i_2 \cdots i_k) \\ [\dot{1}_k & \dot{0}_{n-k}] \\ (j_1 & j_2 \cdots j_k) \end{bmatrix} \middle\| \begin{pmatrix} [m]_n \\ [m]_{n-1} \end{pmatrix} \right\rangle$$

$$= (-1)^{k(k-1)/2} \prod_{l,s=1}^{k} S(j_l - i_s) \left[\prod_{l=1}^{k} \left(\prod_{\substack{j=1 \\ j \neq (j_1 j_2 \cdots j_k)}}^{n} \frac{(p_{i_l n} - p_{j,n-1})}{(p_{j_l,n-1} - p_{j,n-1} + 1)} \prod_{\substack{i=1 \\ i \neq (i_1 i_2 \cdots i_k)}}^{n} \frac{(p_{j_l,n-1} - p_{in} + 1)}{(p_{i_l n} - p_{in})} \right) \right]^{1/2} \tag{2.3}$$

for $i_1 < i_2 < \cdots < i_k$ and $j_1 < j_2 < \cdots j_k$,

where

$$S(j-i) = \begin{cases} +1 & j \geqslant i \\ -1 & j < i, \end{cases} \quad \text{and} \quad p_{in} = m_{in} + n - i.$$

For example, for U(4):

$$\left\langle \begin{matrix} m_1 & m_2 & m_3 & 0 \\ m_1' & m_2' & m_3' \end{matrix} \middle| \begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 \end{matrix} \middle| \begin{matrix} m_1+1 & m_2+1 & m_3 & 0 \\ m_1'+1 & m_2'+1 & m_3' \end{matrix} \right\rangle$$

$$= \left(\frac{(m_1' - m_3 + 2)(m_1' + 3)(m_2 - m_3' + 2)(m_1 - m_3' + 3)(m_2' - m_3 + 1)(m_2' + 2)}{(m_1' - m_3' + 3)(m_1 + 3)(m_2 - m_3 + 1)(m_1 - m_3 + 2)(m_2' - m_3' + 2)(m_2 + 2)} \right)^{1/2}. \tag{2.4}$$

From these results one can also obtain explicit expressions for the Racah coefficients of U(n), involving the elementary representations $[\dot{1}_k, \dot{0}_{n-k}]$ and some simple irreducible representations. For example, in U(4), we find

$$\begin{Bmatrix} [1\ 0\ 0\ 0] & [1\ 0\ 0\ 0] & [1\ 1\ 0\ 0] \\ [m_1\ m_2\ m_3\ 0] & [m_1+1\ m_2+1\ m_3\ 0] & [m_1+1\ m_2\ m_3\ 0] \end{Bmatrix}$$

$$= -\{\dim[1\ 1\ 0\ 0]_{U(4)} \dim[m_1+1\ m_2\ m_3\ 0]_{U(4)}\}^{-1/2} \left\{ \frac{(m_1 - m_2 + 2)}{2(m_1 - m_2 + 1)} \right\}^{1/2}. \tag{2.5}$$

Similarly,

$$\begin{Bmatrix} [1\ 0\ 0\ 0] & [1\ 0\ 0\ 0] & [1\ 1\ 0\ 0] \\ [m_1\ m_2\ m_3\ 0] & [m_1+1\ m_2+1\ m_3\ 0] & [m_1\ m_2+1\ m_3\ 0] \end{Bmatrix}$$

$$= -\{\dim[1\ 1\ 0\ 0]_{U(4)} \dim[m_1\ m_2+1\ m_3\ 0]_{U(4)}\}^{-1/2} \left\{ \frac{(m_1 - m_2)}{2(m_1 - m_2 + 1)} \right\}^{1/2}. \tag{2.6}$$

At present, though the multiplicity-free Racah coefficients of U(n) involving the elementary representation $[\dot{1}_k, \dot{0}_{n-k}]$ can be expressed as a sum over the product of four Wigner coefficients, they have not yet been obtained in a further simplified form. For example, in U(4), one would like to know whether

$$\begin{Bmatrix} [1,1,0,0] & [m_1,m_2,m_3,0] & [m_1+1,m_2+1,m_3,0] \\ [1,1,0,0] & [m_1+2,m_2+2,m_3,0] & [m_1,m_2+1,m_3+1,0] \end{Bmatrix}$$

can be expressed in a simpler form, other than the product of four Wigner coefficients summed up. We have good reason to believe that this can be done, using the techniques developed in I and II. However, we shall leave this topic to be discussed in detail in a future publication.

The Wigner and Racah coefficients of U(n) involving $[p,\dot{0}]$ and $[\dot{p},0]$ have been treated in I and II and explicitly simple and closed expressions for these coefficients have been given there.

III. PHASE FACTORS FOR THE MULTIPLICITY-FREE WIGNER COEFFICIENTS OF U(N)

In I, we gave the phase convention as follows: We extend the Condon–Shortley phase convention from U(2) to U(n) by the following mapping:

$$j \to \epsilon_n (m_{1n} - m_{nn}), \tag{3.1}$$

$$m \to \tfrac{1}{2} \sum_{i=1}^{n-1} z_i, \tag{3.2}$$

where ϵ_n and z_i are suitably defined quantities. However, we find that for $n \geqslant 4$, we have to modify (3.1). This becomes clear if we are to write down the most general requirements that the phase convention must satisfy. We list the following seven conditions as the most general ones that the phase convention must satisfy.

(1) Under conjugation,, $j \ast \rightarrow j$.

(2) Under conjugation, $m^{\ast} \rightarrow -m$.

(3) The phase factor occurring in all the symmetry relations must be an integer.

(4) The mapping of j must contain m_{in} only for U(n).

(5) m must be mapped to diagonal operators.

(6) It is desirable to keep the phase convention in agreement with Baird and Biedenharn,[7] which essentially maps $j - m$ to $\rho_{n-1} - \rho_{n-1(\max)}$, where $\rho(n) = \Sigma_{j=1}^{n} \Sigma_{i=1}^{j} m_{ij}$.

(7) The phase convention must agree with the formula for the multiplicity-free Wigner coefficients of U(n) given by Chacón, Ciftan, and Biedenharn[8] (denoted by CCB) or Alisauskas, Jucys, and Jucys[9] (denoted by AJJ). We shall show below that this is equivalent to the following statement: For the state

$$\left| \begin{matrix} p, & & \dot{0} \\ q, & & \dot{0} \\ & \dot{0} & \end{matrix} \right\rangle$$

in U(n), $j + m \rightarrow q$.

Finally, it is understood, of course, that the mapping of j and m is an extension of the phase convention from U(2) to U(n).

We shall now prove the statement given in condition (7), i.e., the formula for the multiplicity-free Wigner coefficients of U(n) as given by CCB or AJJ requires that $j + m$ be mapped to q for the state

$$\left| \begin{matrix} p, & & \dot{0} \\ q, & & \dot{0} \\ & \dot{0} & \end{matrix} \right\rangle.$$

(I wish to take this opportunity here to thank Dr. Wayne J. Holman III for suggesting to me to look into this condition.)

First of all, in U(2), we have the following symmetry:

$$C^{j_1 j_2 j}_{m_1 m_2 m} = (-1)^{j_2 + m_2} \left(\frac{(2j+1)}{(2j_1+1)} \right)^{1/2} C^{j j_2 j_1}_{-m\, m_2\, -m_1}.$$

So what we are looking for is a similar relation in U(n):

$$\left\langle (M) \left| \left\langle \begin{matrix} p, & & \dot{0} \\ q, & & \dot{0} \\ & \dot{0} & \end{matrix} \right| \right| (m) \right\rangle = x \left\langle (m)^{\ast} \left| \left\langle \begin{matrix} p, & & \dot{0} \\ q, & & \dot{0} \\ & \dot{0} & \end{matrix} \right| \right| (M)^{\ast} \right\rangle, \tag{3.3}$$

where x is to be determined from the CCB or AJJ formula. We take

$$|(m)\rangle = \left| \begin{matrix} h'_1 \cdots h'_n \\ q'_1 \cdots q'_{n-1} \\ \max \end{matrix} \right\rangle \quad \text{and} \quad |(M)\rangle = \left| \begin{matrix} h_1 \cdots h_n \\ q_1 \cdots q_{n-1} \\ q'_1 \cdots q'_{n-1} \\ \max \end{matrix} \right\rangle,$$

where, of course,

$$\sum_{i=1}^{n} h'_i + p = \sum_{i=1}^{n} h_i, \quad \sum_{i=1}^{n-1} q' + q = \sum_{i=1}^{n-1} q_i.$$

We then find that

$$|(m)^{\ast}\rangle = \left| \begin{matrix} h'_1 - h'_n \cdots h'_1 - h'_2 & 0 \\ h'_1 - q'_{n-1} \cdots h'_1 - q'_1 & \\ h'_1 - q'_{n-2} \cdots h'_1 - q'_1 & \end{matrix} \right\rangle, \quad |(M)^{\ast}\rangle = \left| \begin{matrix} h'_1 - h_n \cdots h'_1 - h_1 \\ h'_1 - q_{n-1} \cdots h'_1 - q_1 \\ h'_1 - q'_{n-2} \cdots h'_1 - q'_1 \end{matrix} \right\rangle$$

Now we apply the CCB formula to both sides of (3.3) and find that

$$x = (-1)^{q} \left(\frac{\dim(M)}{\dim(m)} \right)^{1/2}. \tag{3.4}$$

This can be seen as follows. Applying the CCB formula to the left side of (3.3), we obtain a factor times a summation over $n - 1$ variables. Let us call these summation variables ρ_i, $i = 1,\ldots,n-1$. The right side also contains a factor times a summation over $n - 1$ variables. Let us call these ρ'_i, $i = 1,\ldots,n-1$. If we set

$$\rho'_i = q_{n-i} - q'_{n-i} - \rho_{n-i}, \tag{3.5}$$

we find that the two terms under the summation signs are equal, except for the phase factor $(-1)^q$. The other two factors then contribute to the dimensional factor $(\dim(M)/\dim(m))^{1/2}$. This proves our statement that

$j+m\to q$ for the state $\left|\begin{matrix} p, & & \dot{0} \\ q, & & \dot{0} \\ & \dot{0} & \end{matrix}\right\rangle$ in U(n).

We now assert that the following phase convention satisfies all the seven requirements listed above.

$$\text{(a) } m\to \tfrac{1}{2}\sum_{i=1}^{n-1} z_i, \quad \text{where } z_i=(i+1)\sum_{j=1}^{i} m_{ji} - i\sum_{j=1}^{i+1} m_{j,i+1}. \tag{3.6}$$

(b) For U($4k$), $j\to(3/2)(m_{1n}-m_{2n}+m_{3n}-m_{4n}\cdots+m_{4k-1,n}-m_{nn})$.

$$\begin{aligned} &\text{For U}(4k+1), j\to m_{2n}+m_{4n}+m_{6n}\cdots+m_{4k,n}. \\ &\text{For U}(4k+2), j\to \tfrac{1}{2}(m_{1n}-m_{2n}+m_{3n}-m_{4n}+m_{4k+1,n}-m_{nn}). \\ &\text{For U}(4k+3), j\to m_{1n}+m_{3n}+m_{5n}\cdots+m_{nn}. \end{aligned} \tag{3.7}$$

We shall omit the proof that the phase convention given above satisfies all the seven requirements, since it is an elementary calculation.

We now wish to apply this phase convention to a special case which was found to be very important in the evaluation of the multiplicity-free Wigner and Racah coefficients of U(n),

$$\left\langle \begin{matrix} [a,0] & [b]_n \\ (a)_{n-1} & (\beta)_{n-1} \\ \max & \end{matrix} \middle| \begin{matrix} [j]_n \\ (\theta)_{n-1} \end{matrix} \right\rangle = (-1)^y \left(\frac{\dim[j]_n}{\dim[b]_n}\right)^{1/2} \left\langle \begin{matrix} [j]_n & [a,\dot{0}] \\ (\theta)_{n-1} & (\dot{0}) \end{matrix} \middle| \begin{matrix} [b]_n \\ (\beta)_{n-1} \end{matrix} \right\rangle. \tag{3.8}$$

In I, we started that $y=0$ for U(2), U(3),...,U($2k+1$), but need not be zero for U($2k+2$), $k=1,2,\cdots$. We wish to correct this statement and say that $y=0$ for all U(n). This is easy to see. Equation (3.8) is the generalization of Eq. (3.5.16) of Edmonds.[10] Thus the phase factor y should be that corresponding to j_1-m_1, i.e., according to condition (6),

$$\rho_{(n-1)} - \rho_{\substack{(n-1) \\ \max}} = 0.$$

The result is that in Eq. (1.17) of I, the phase factor should always be zero.

Incidentally, there is a misprint in the dimensional factor of Eq. (1.17) of I. The correct expression should therefore read:

$$\begin{aligned} &\left\langle \begin{matrix} m_{1n}\cdots m_{n-1,n}, & 0 \\ m_{1n-1}\cdots m_{n-1n-1} & \end{matrix} \middle| \begin{matrix} p, & \dot{0} \\ q, & \dot{0} \end{matrix} \middle| \begin{matrix} m'_{1n}\cdots m'_{n-1n}, & 0 \\ m'_{1n-1}\cdots m'_{n-1n-1} & \end{matrix} \right\rangle \\ &= \left(\frac{\mathscr{M}(m'_{1n}\cdots m'_{n-1n})W_n!\,\mathscr{M}(m_{1n-1}\cdots m_{n-1n-1})p!}{\mathscr{M}(m_{1n}\cdots m_{n-1n})W'_n!\,\mathscr{M}(m'_{1n-1}\cdots m'_{n-1n-1})(p-q)!q!}\right)^{1/2} \\ &\quad\times[\dim(m'_{1n}\cdots m'_{n-1n})\dim(m_{1n-1}\cdots m_{n-1n-1})\dim(W_n,\dot{0})_{n-1}\dim(p,\dot{0})_{n-1}]^{1/2} \\ &\quad\times X\begin{vmatrix} [\dot{q},0]_{n-1} & [m_{1n-1}\cdots m_{n-1n-1}] & [m'_{1n-1}\cdots m'_{n-1n-1}] \\ [p-q,0]_{n-1} & [W_n,\dot{0}_{n-1}] & [W'_n,\dot{0}] \\ [\dot{p},0]_{n-1} & [m_{1n}\cdots m_{n-1n}] & [m'_{1n}\cdots m'_{n-1n}] \end{vmatrix}. \end{aligned} \tag{3.9}$$

This change will also affect the phase factors for the multiplicity-free Racah coefficients of U(n) discussed in II. We shall discuss this in the next section.

We conclude this section by stating that all the symmetry properties for the multiplicity-free Wigner coefficients of U(n) can be directly taken over from the corresponding symmetry properties for U(2) or R (3), by the mapping (3.6) and (3.7).

IV. PHASE FACTORS FOR THE MULTIPLICITY-FREE RACAH COEFFICIENTS OF U(N)

In II we have obtained multiplicity-free Racah coefficients of U(n) where one of the columns in the 6-j symbol consists of 1. $(\dot{p},0)$, $(\dot{q},0)$, 2. $(\dot{p},0)$, $(q,\dot{0})$ and 3. $(p,\dot{0})$, $(q,\dot{0})$. However, some of the phase factors given in II will have to be corrected. We first give the correct expressions, and then give the proof later on.

Equations (2.6)–(2.12) in II are correct, except for a misprint in Eq. (2.7). It should read

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j^*_{12} & j_2 & j^*_1 \\ j^*_{23} & J^* & j^*_3 \end{Bmatrix}. \tag{4.1}$$

Equation (3.3) of II should read

$$X\begin{vmatrix} [m'_{1n-1}\cdots m'_{n-1n-1}] & [q,\dot{0}]_{n-1} & [m_{1n-1}\cdots m_{n-1n-1}] \\ [W'_n,\dot{0}]_{n-1} & [p-q,\dot{0}]_{n-1} & [W_n,\dot{0}]_{n-1} \\ [m'_{1n}\cdots m'_{n-1n}] & [p,\dot{0}]_{n-1} & [m_{1n}\cdots m_{n-1n}] \end{vmatrix}$$

$$= \{\dim[W_n,\dot{0}]_{n-1}\dim[p,\dot{0}]_{n-1}\dim[m'_{1n}\cdots m'_{n-1,n}]_{U(n-1)}\dim[m_{1n-1}\cdots m_{n-1n-1}]\}^{1/2}$$

$$\times\left\{\frac{\mathscr{M}(m_{1n}\cdots m_{n-1n})\mathscr{M}(m'_{1n-1}\cdots m'_{n-1n-1})}{\mathscr{M}(m'_{1n}\cdots m'_{n-1n})\mathscr{M}(m_{1n-1}\cdots m_{n-1n-1})}\frac{W'_n!(p-q)!q!}{W_n!p!}\right\}^{1/2}$$

$$\times\left\langle \begin{matrix} m_{1n}\cdots m_{n-1n} & 0 \\ m_{1n-1}\cdots m_{n-1n-1} & \end{matrix}\middle|\middle| \begin{matrix} p, & \dot{0} \\ q, & \dot{0} \end{matrix}\middle|\middle| \begin{matrix} m'_{1n}\cdots m'_{n-1n} & 0 \\ m'_{1n-1}\cdots m'_{n-1n-1} & \end{matrix}\right\rangle. \tag{4.2}$$

If one interchanges columns 1 and 2 in the 9-j symbol on the left of Eq. (4.2), one should multiply by a phase factor $(-1)^{\Sigma \text{ all the } 9\text{-}j\text{'s}}$. This phase factor can be explicitly calculated according to the mapping of (3.7). In what follows, we shall denote the mapping of j by $j \to \epsilon_n[m]_n$ for U(n), according to (3.7).

Equation (3.4) of II should read

$$\begin{Bmatrix} [W_{n+1},\dot{0}] & [m_{1n-1}\cdots m_{n-1n-1},\dot{0}] & [m'_{1n}\cdots m'_{nn}] \\ [p,\dot{0}] & [m_{1n+1}\ldots m_{nn+1}] & [m_{1n}\cdots m_{nn}] \end{Bmatrix}$$

$$= (-1)^{\epsilon_n([m'_n]+[p]+[W_{n-1}]+[m_n])}[\dim(m'_{1n}\cdots m'_{nn})\dim(m_{1n}\cdots m_{nn})]^{1/2}$$

$$\times\left(\frac{\mathscr{M}(m_{1n+1}\cdots m_{nn+1})\mathscr{M}(m_{1n-1}\cdots m_{n-1n-1}0)}{\mathscr{M}(m'_{1n}\cdots m'_{nn})\mathscr{M}(m_{1n}\cdots m_{nn})}\right)^{1/2}\left\langle \begin{matrix} m_{1n+1}\cdots m_{nn+1} & 0 \\ m_{1n}\cdots m_{nn} & \end{matrix}\middle|\middle| \begin{matrix} p, & \dot{0} \\ p, & \dot{0} \end{matrix}\middle|\middle| \begin{matrix} m'_{1n}\cdots m'_{nn} & 0 \\ m_{1n-1}\cdots m_{n-1n-1} & 0 \end{matrix}\right\rangle_{U(n+1)}. \tag{4.3}$$

Finally Eq. (3.7) or II should read

$$\begin{Bmatrix} [W_{n+1},\dot{0}] & [m_{1n-1}\cdots m_{n-1n-1}0] & [m'_{1n}\cdots m'_{nn}] \\ [p,\dot{0}] & [m_{1n+1}\cdots m_{nn+1}] & [m_{1n}\cdots m_{nn}] \end{Bmatrix}$$

$$= \begin{Bmatrix} [W_{n-1},\dot{0}] & [m_{1n}\cdots m_{nn}] & [m_{1n+1}\cdots m_{nn+1}] \\ [p,\dot{0}] & [m'_{1n}+p,\cdots m'_{nn}+p] & [m_{1n-1}+p,\ldots,m_{n-1n-1}+p,p] \end{Bmatrix}. \tag{4.4}$$

We now prove (4.4) by using two different methods. The first method makes use of Eqs. (2.6)–(2.12) of II exclusively. This proof is due to Holman, and I wish to thank him for pointing this out to me as well as for urging me to reexamine the whole question of phase factors for the multiplicity-free Wigner and Racah coefficients of U(n). The second method uses Eqs. (1.17) of I [equivalent to Eq. (3.9) here], and Eqs. (3.3) and (3.4) of II [equivalent to Eqs. (4.2) and (4.3) here].

First proof: Using simplified notations we wish to prove

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} = \begin{Bmatrix} j_1 & j_{23} & J \\ j_3^* & j_{12} & j_2 \end{Bmatrix}.$$

Interchange columns 1 and 2, using (2.6) of II,

$$\begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} \to \begin{Bmatrix} j_2 & j_1 & j_{12} \\ J^* & j_3^* & j_{23}^* \end{Bmatrix}.$$

Then interchange columns 1 and 3, using (2.7) of II,

$$\to \begin{Bmatrix} j_{12}^* & j_1 & j_2^* \\ j_{23}^* & j_3 & J \end{Bmatrix}.$$

Now invert columns 1 and 3, using (2.9) of II,

$$\to \begin{Bmatrix} j_{23}^* & j_1^* & J^* \\ j_{12} & j_3^* & j_2 \end{Bmatrix}.$$

Now interchange columns 1 and 2, using (2.6) of II,

$$\to \begin{Bmatrix} j_1^* & j_{23}^* & J^* \\ j_3 & j_{12}^* & j_1^* \end{Bmatrix}.$$

Now conjugate using (2.10) of II,

$$\to \begin{Bmatrix} j_1 & j_{23} & J \\ j_3^* & j_{12} & j_2 \end{Bmatrix}.$$

Q.E.D.

Second proof: Put $q = p$ in Eq. (1.17) of I [equivalent to Eq. (3.8) here]. We have:

$$X \begin{vmatrix} [\dot{p},0] & [m]_n & [m']_{n-1} \\ [\dot{0}] & [W_{n+1},\dot{0}] & [W_{n+1},\dot{0}] \\ [\dot{p},0] & [m]_{n+1} & [m']_n \end{vmatrix}_{U(n)}$$

$$= (-1)^{\epsilon_n([p]+[W_{n-1}]+[m'_n]+[m_n])}\{\dim[p,\dot{0}]_{U(n)} \dim[W_{n+1},\dot{0}]_{U(n)}\}^{-1/2} \begin{Bmatrix} [W_{n+1},\dot{0}] & [m]_n & [m]_{n+1} \\ [p,\dot{0}] & [m']_n & [m']_{n-1} \end{Bmatrix}_{U(n)}$$

$$= \left\langle \begin{matrix} m_{1n+1}\cdots m_{nn+1} & 0 \\ m_{1n}\cdots m_{nn} \end{matrix} \middle| \begin{matrix} p, & \dot{0} \\ p, & \dot{0} \end{matrix} \middle| \begin{matrix} m'_{1n}\cdots m'_{nn} & 0 \\ m'_{1n-1}\cdots m'_{n-1n-1} & 0 \end{matrix} \right\rangle_{U(n+1)} \{\dim[p,\dot{0}]_{U(n)} \dim[W_{n+1},\dot{0}]_{U(n)}\}^{-1/2}$$

$$\times \{\dim[m']_n \dim[m]_n\}^{-1/2} \times mf, \tag{4.5}$$

where mf = inverse of the measure on the right of Eq. (3.8)

$$= \left\{\frac{\mathscr{M}[m]_{n+1}\mathscr{M}[m']_{n-1}}{\mathscr{M}[m']_n \mathscr{M}[m]_n}\right\}^{1/2}. \tag{4.6}$$

On the other hand we have, from Eq. (3.4) of II [equivalent to Eq. (4.3) here],

$$\begin{bmatrix} [W_{n+1},\dot{0}] & [m']_{n-1} & [m']_n \\ [p,\dot{0}] & [m]_{n+1} & [m]_n \end{bmatrix} = (-1)^{\epsilon_n([p]+[W_{n-1}]+[m'_n]+[m_n])}\{\dim[m']_n \dim[m]_n\}^{-1/2} \times mf$$

$$\times \left\langle \begin{matrix} m_{1n+1}\cdots m_{nn+1} & 0 \\ m_{1n}\cdots m_{nn} \end{matrix} \middle| \begin{matrix} p, & \dot{0} \\ p, & \dot{0} \end{matrix} \middle| \begin{matrix} m'_{1n}\cdots m'_{nn} & 0 \\ m'_{1n-1}\cdots m_{n-1n-1} & 0 \end{matrix} \right\rangle. \tag{4.7}$$

Comparing (4.5) and (4.7) we obtain the desired result.

ACKNOWLEDGMENT

I wish to thank Dr. Wayne J. Holman III for urging me to reexamine the phase factors and for many helpful and valuable discussions and suggestions.

APPENDIX

In this appendix we wish to prove the following statement. The outer multiplicity of $[m_3]$, $\tilde{y}(m_3)$, cannot be greater than $y(M)$, where $y(M)$ is the highest inner multiplicity of a certain weight M belonging to $[m_2]$. The proof follows.

(1) Write the irreducible representation as $[m_{n1},m_{n2},\ldots,m_{n,n-1},0]$ for U(n), with $m_{n1} \geqslant m_{n2} \geqslant m_{n3} \cdots \geqslant m_{n,n-1}$.

(2) Arrange the weights of the irreducible representation according to dominance. For example, (3,2,0,0) > (3,1,1,0). The weights are also written in such a way that the last number is 0.

(3) From Racah's recurrence relation, it can be proved that $y(w_1) \geqslant y(w_2)$, if $w_1 < w_2$, where the second inequality refers to dominance, i.e., w_2 is more dominant than w_1. (For proof, see Wong and Gruber, J. Math. Phys. **11**, 3187, 1970.)

(4) From the Racah–Speiser lemma, we now calculate the outer multiplicity for a given irreducible representation $[m_3]$ as follows. First take $S = 1$, the identity operation. The corresponding term must be the inner multiplicity of a certain weight, say w_0, belonging to $[m_2]$. Then take the Weyl operation consisting of a 2-transposition, e.g., $\left(\begin{smallmatrix} 2 & 3 & 4 \\ 3 & 4 & 2 \end{smallmatrix}\right) = (2\,4)(2\,3)$. The resulting term from this permutation will have a + sign. However, from the formula of the Racah–Speiser lemma, Eq. (2.1), the resulting weight belonging to $[m_2]$ must be at least (3,3,0), or (3,0,0) if we have the 2-transposition $\left(\begin{smallmatrix} 2 & 3 & 4 \\ 4 & 2 & 3 \end{smallmatrix}\right)$. We claim that for every 2-transposition (which will involve at least three positions, e.g., 2,3,4), there is at least one 1-transposition resulting from the same positions (i.e., transposition of two out of the three positions mentioned above) which will contribute a nonvanishing term in Eq. (2.1).

This 1-transposition will give a − sign. Moreover, $y(w_1) \geqslant y(w_2)$, because $w_1 < w_2$, since w_1 results from a 1-transposition while w_2 results from a 2-transposition arising from the same positions. Similar arguments can be applied to higher transpositions. In other words, for every $2k$-transposition, there exists at least one $(2k-1)$-transposition consisting of the same terms in the application of the Racah–Speiser lemma. Moreover, $y(w_{2k-1}) \geqslant y(w_{2k})$ because $w_{2k-1} < w_{2k}$. Thus we have

$$\tilde{y}(m_3) = y(w_0) - [y(w_1 - y(w_2)] - [y(w_3) - y(w_4)] - \cdots - [y(w_{2k-1}) - y(w_{2k})] \cdots,$$

where all terms in the brackets are nonnegative. Since the Weyl group of U(n) consists of permutations only, and since all permutations can be reduced to transpositions, this completes the proof.

[1] J.R. Derome, J. Math. Phys. **7**, 612 (1966).
[2] J.R. Derome, J. Math. Phys. **8**, 714 (1967).
[3] P.H. Butler, Philos. Trans. R. Soc. London, Ser. A **277**, 545 (1975), and references thereof.
[4] M.K.F. Wong, J. Math. Phys. **17**, 1558 (1976), henceforth referred to as I.
[5] M.K.F. Wong, J. Math. Phys. **19**, 1635 (1978), henceforth referred to as II.
[6] L.C. Biedenharn and J.D. Louck, Commun. Math. Phys. **8**, 89 (1968).
[7] G.E. Baird and L.C. Biedenharn, J. Math. Phys. **5**, 1723 (1964).
[8] E. Chacón, M. Ciftan, and L.C. Biedenharn, J. Math. Phys. **13**, 577 (1972).
[9] S.J. Ališauskas, A.A.A. Jucys, and A.P. Jucys, J. Math. Phys. **13**, 1329 (1972).
[10] A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1960).

On the definition and properties of generalized 6-j symbols

Jacques Raynal

CEN de Saclay, Boite Postale No. 2–91190 Gif-Sur-Yvette, France
(Received 10 July 1978; revised manuscript received 13 November 1978)

Whipple's work on the symmetries of well-poised ${}_7F_6(1)$ and Saalschützian ${}_4F_3(1)$ series with unit argument is applied to study the properties of 6-j symbols generalized to any arguments. For SU2, we obtain eleven different looking ${}_4F_3(1)$ series which can be used. Whipple's parameter x s provide a good description of symmetries. We obtain quite simple recurrence relations, valid for any arguments, in terms of these parameters.

I. INTRODUCTION

In a previous paper[1] we discussed the different formulas which can be used for 3-j symbols of SU2. We generalized one of the formulas to obtain a generalized hypergeometric function ${}_3F_2$ with unit argument and complex parameters. Whipple[2] showed that there are twelve sets of ten such functions which are equivalent, with relations between any three of these sets. We studied the consequences of Whipple's relations when all the parameters become integer. In so doing, we determined that any of the 96 finite ${}_3F_2$ can be used for SU2 or SU(1,1), for which all the parameters are integer, but also, that the 56 finite ${}_3F_2$ can be used in the hyperspherical formalism[3] for which some of the parameters are half-integer. These half-integer parameters are replaced by any value when relations between Laguerre and Jacobi polynomials[4] are studied independently of group theory.

The 6-j symbols appear in these problems as permutation coefficients. They are Saalschützian generalized hypergeometric series ${}_4F_3$ with unit argument. The parameters are integer for SU2 and SU(1,1); some of them are half-integer in the hyperspherical formalism[3,5] and can take any value for relations between Jacobi polynomials.[6] As for 3-j symbols, there are different formulas[6,7] and the problem of identification between results of different authors arises.

Relations between Saalschützian ${}_4F_3$ series with unit argument are known only for terminating series; using them, we cannot generalize all the parameters. However, for any parameters, there are relations between some special well-poised ${}_7F_6$ series of argument unity which can be expressed as sums of two Saalschützian ${}_4F_3$, one of which disappears when the other terminates. These relations have been also studied by Whipple.[8] They have been used to study symmetries and relations between 6-j symbols of SU(2) and SU(1,1) where all the parameters of the generalized hypergeometric series are integer, by D'Adda *et al.*[9] There are 12 sets of 16 well-poised ${}_7F_6$ series which are equivalent, with relations between any three of these sets. To any ${}_7F_6(1)$ series, 120 generalized 6-j symbols can be associated by permutation of the free parameters. There are 160 Saalschützian ${}_4F_3(1)$ series involved, each of them related to 144 6-j symbols by permutation of its parameters. These symmetries which involve $144\times160 = 16\times12\times120 = 23040$ elements include Yutsis' mirror symmetry[10]; without "mirror symmetry" there are only 144 elements, which are those of the usual and Regge's symmetry,[11] as summarized by Shelepin's symbol.[12]

In Sec. 2 we shall summarize Whipple's work. In Sec. 3, we choose a definition for a generalized 6-j symbol which reduces to a usual formula when the arguments fulfill the usual relations. We list the different ${}_4F_3(1)$ series obtained for SU2, taking into account usual symmetries but not Regge's ones, which are not seen directly on the parameters. In Sec. 4, the existence of a negative integer among parameters leads to other formulas. The consequences of the existence of more than one negative integer are studied in Sec. 5. Then we discuss the symmetry properties and the analytical continuation in Sec. 6. This was exactly the plan of the work on 3-j symbols.[1] However, in Sec. 7 we derive recurrence relations from the expression in terms of ${}_7F_6(1)$ series and in an Appendix we give similar expressions for 3-j symbols.

2. RELATIONS BETWEEN SAALSCHÜTZIAN ${}_4F_3$

For a Saalschützian generalized hypergeometric series ${}_4F_3[A, B, C, D; E, F, G; 1]$ (Saalschützian means that the convergency indicator $s = E + F + G - A - B - C - D$ is unity) Whipple defines 6 parameters x such that

$$A = \tfrac{1}{2} - x_0 - x_1 - x_2 - x_3 - x_4 - x_5, \quad B = \tfrac{1}{2} - x_0 - x_1 - x_2 - x_3 + x_4 + x_5, C = \tfrac{1}{2} - x_0 - x_1 - x_2 + x_3 + x_4 + x_5,$$
$$D = \tfrac{1}{2} - x_0 - x_1 - x_2 + x_3 + x_4 - x_5, \quad E = 1 - 2x_1 - 2x_2, \quad F = 1 - 2x_0 - 2x_2, \quad G = 1 - 2x_0 - 2x_1. \tag{1}$$

Whipple's parameters are

$$x_0 = \tfrac{1}{4}(1 + E - F - G), \quad x_1 = \tfrac{1}{4}(1 - E + F - G), \quad x_2 = \tfrac{1}{4}(1 - E - F + G),$$
$$x_3 = \tfrac{1}{4}(-A - B + C + D), \quad x_4 = \tfrac{1}{4}(-A + B - C + D), \quad x_5 = \tfrac{1}{4}(-A + B + C - D). \tag{2}$$

 0022-2488/79/122398-18$01.00

Due to invariance under permutation of (E,F,G) and (A,B,C,D), these parameters are defined up to a permutation of (x_0,x_1,x_2) and (x_3,x_4,x_5) plus an even number of changes of sign among (x_3,x_4,x_5).

The set of Saalschützian ${}_4F_3(1)$ series between which Whipple studied the relations, is characterized by the permutations of $(x_0\cdots x_5)$ and the change of sign for an even number among them. These operations define $2^5\times 6! = 23040$ sets of parameters corresponding to 160 different ${}_4F_3(1)$ series. To simplify notation of numerator parameters we shall use $(i,j,k,l,m,n$ all different)

$$\omega_{ij}^{klmn} = \omega_{ij} = \omega^{klmn} = \tfrac{1}{2} - x_i - x_j + x_k + x_l + x_m + x_n, \tag{3}$$

$$\omega_{ijklmn} = \omega_- = \tfrac{1}{2} - x_i - x_j - x_k - x_l - x_m - x_n, \quad \omega^{ijklmn} = \omega^+ = \tfrac{1}{2} + x_i + x_j + x_k + x_l + x_m + x_n,$$

that is, upper indices for the x's with plus sign and lower indices for the x's with minus sign: one set of indices can be dropped and $+$ or $-$ means that all the x's appear with plus or minus sign.

Instead of the ${}_4F_3$, Whipple considers

$$S(i,j,k) = \frac{1}{\Gamma(1-2x_i-2x_j)\Gamma(1-2x_j-2x_k)\Gamma(1-2x_k-2x_i)}$$

$$\times {}_4F_3[\omega_-,\omega^{lm},\omega^{ln},\omega^{mn};1-2x_i-2x_j,1-2x_j-2x_k,1-2x_k-2x_i;1], \tag{4}$$

and similar expressions for $S(\pm i,\pm j,\pm k)$ with a change of sign for the corresponding x's keeping the rule of an even number of minus signs in the numerator parameters. The definitions (1) and (2) are related to $S(0,1,2)$. In the general case, there are relations between four S's. However, the function

$$\mathcal{G}_p(0) = \frac{\pi}{\sin 2\pi(x_1+x_2)}\left(\frac{S(0,1,2)}{\Gamma(\omega^{12})\Gamma(\omega_{03})\Gamma(\omega_{04})\Gamma(\omega_{05})} - \frac{S(0,-1,-2)}{\Gamma(\omega_-)\Gamma(\omega^{34})\Gamma(\omega^{35})\Gamma(\omega^{45})}\right)$$

$$= \frac{\Gamma(1+a)}{\Gamma(\omega_-)\Pi_{i=1}^5\,[\Gamma(1+a-\omega_{0i})]}\,{}_7F_6\left[a,\tfrac{1}{2}a,(\omega_{0i});1+\tfrac{1}{2}a,(1+a-\omega_{0i});1\right], \tag{5}$$

where $a = \omega^+ - 2x_0$ and $i = 1$ to 5, is invariant by permutation of $(x_1\cdots x_5)$ which does not change the parameters of the ${}_7F_6(1)$ series, which is a well poised generalized hypergeometric series with a special value of the second parameter. Moreover, as $\mathcal{G}_p(0)$, as defined in terms of S's, is invariant for an even number of changes of sign among (x_3,x_4,x_5), it is also invariant for an even number of changes of sign among $(x_1\cdots x_5)$ which generates sixteen different ${}_7F_6(1)$ series. Note that the arguments of Γ functions which divide the S's functions are the numerator parameters of the other one; if one of them is a negative integer, the finite ${}_4F_3(1)$ series remains alone.

In such a way, we can define six $\mathcal{G}_p(i)$ for $i = 0$–6 and six $\mathcal{G}_n(i)$ which differ from $\mathcal{G}_p(i)$ by a change of sign for x_i. Each $\mathcal{G}_p$ or $\mathcal{G}_n$ can be expressed in 16 different ways by a well-poised ${}_7F_6(1)$ series and 20 different ways as a sum of two Saalschützian ${}_4F_3(1)$ series. Whipple obtains

$$\frac{\sin 2\pi(x_1+x_0)\mathcal{G}_p(0)}{\Gamma(\omega^{02})\Gamma(\omega_{13})\Gamma(\omega_{14})\Gamma(\omega_{15})} - \frac{\sin 2\pi(x_0+x_2)\mathcal{G}_p(1)}{\Gamma(\omega^{12})\Gamma(\omega_{03})\Gamma(\omega_{04})\Gamma(\omega_{05})} = \frac{\sin 2\pi(x_1-x_0)\mathcal{G}_n(2)}{\Gamma(\omega_-)\Gamma(\omega^{34})\Gamma(\omega^{35})\Gamma(\omega^{45})} \tag{6}$$

and deduces a relation between $\mathcal{G}_p(0),\mathcal{G}_p(1)$, and $\mathcal{G}_p(3)$ from (6) and the relation obtained by exchange of 0 and 3 in (6). From there he obtains a relation between any three of these functions.

3. GENERALIZED 6-j SYMBOLS

A. Definitions

For the definition of a generalized 6-j symbol, we choose

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)]\,\frac{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)}{\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(b-d+f+1)\Gamma(c-d+e+1)}$$

$$\times\left(\frac{\Gamma(b+c+e+f+2)}{\Gamma(a-b+d-e+1)\Gamma(a-c+d-f+1)}\,{}_4F_3[a-e-f,a-b-c,-c+d-e,-b+d-f;\right.$$

$$\left.-b-c-e-f-1,a-c+d-f+1,a-b+d-e+1;1]\right.$$

$$+\frac{\Gamma(-b-c-e-f-2)\Gamma(a+b+c+2)\Gamma(a+e+f+2)\Gamma(b+d+f+2)}{\Gamma(a+b+d+e+3)\Gamma(a+c+d+f+3)\Gamma(a-b-c)\Gamma(a-e-f)\Gamma(-b+d-f)}$$

$$\times\frac{\Gamma(c+d+e+2)}{\Gamma(-c+d-e)}\,{}_4F_3[a+b+c+2,a+e+f+2,b+d+f+2,c+d+e+2;$$

$$b+c+e+f+3,a+b+d+e+3,a+c+d+f+3;1]\Big), \tag{7}$$

where

$$\Delta(a,b,c)=\left(\frac{\Gamma(a+b-c+1)\Gamma(a-b+c+1)\Gamma(-a+b+c+1)}{\Gamma(a+b+c+2)}\right)^{1/2}, \tag{8}$$

for any complex value of a,b,c,d,e, and f. However, the real part of the argument of the Γ functions in the square root (8) must be positive in order to define this square root as the analytical continuation to the positive values when the imaginary part of a,b,c,d,e,f vanishes. The two ${}_4F_3(1)$ series are convergent because they are Saalschützian, i.e., their convergency indicator is unity. But, if $b+c+e+f$ is not an integer and one of $(a-b-c, a-e-f, -b+d-f, -c+d-e)$ is an integer (which must be zero or negative with the limitations stated above) the second ${}_4F_3(1)$ disappears; this also happens when $b+c+e+f$ and two values of $(a-b-c, a-e-f, -b+d-f, -c+d-e)$ are integers. The remaining ${}_4F_3(1)$ reduces to Rose's formula[13] for usual values of a,b,c,d,e, and f.

Identifying the first ${}_4F_3(1)$ with the one of $S(0,-1,-2)$ in (5), Whipple's parameters are

$$\begin{aligned} &x_0=-\tfrac{1}{2}(a+d+1), \quad x_1=-\tfrac{1}{2}(b+e+1), \quad x_2=-\tfrac{1}{2}(c+f+1),\\ &x_3=-\tfrac{1}{2}(a-d), \qquad x_4=-\tfrac{1}{2}(b-e), \qquad x_5=-\tfrac{1}{2}(c-f), \end{aligned} \tag{9}$$

and, with notation (3), Shelepin's symbol[12] is

$$\begin{vmatrix} -c+d+e & b+d-f & a+e-f & a+b-c \\ -b+d+f & c+d-e & a-b+c & a-e+f \\ -a+e+f & -a+b+c & c-d+e & b-d+f \\ \multicolumn{4}{c}{\cdots\cdots} \\ -a-b-c-1 & -a-e-f-1 & -b-d-f-1 & -c-d-e-1 \end{vmatrix} = \begin{vmatrix} -\omega^{01} & -\omega_{23} & -\omega_{24} & -\omega_{25} \\ -\omega^{02} & -\omega_{13} & -\omega_{14} & -\omega_{15} \\ -\omega^{12} & -\omega_{03} & -\omega_{04} & -\omega_{05} \\ \multicolumn{4}{c}{\cdots\cdots} \\ \omega^{+} & \omega_{45} & \omega_{35} & \omega_{34} \end{vmatrix}. \tag{10}$$

Here, we added a last line in such a way that the sum of all the diagonals obtained by circular permutation of columns is -1. This property is conserved in any permutation rows or columns. In the following we shall use i,j,k for any of the indices 0,1,2 and l,m,n for any one of 3,4,5. Limiting the notation to two indices ($\omega^{+},\omega_{-},\omega^{\alpha\beta}$ and $\omega_{\alpha\beta}$) the 16 negativelike ω's are those displayed in formula (10): There are three ω^{ij}, nine ω_{il} which are regular components of Shelepin's symbol and ω^{+} and three ω_{lm} which play a different rôle.

With this notation

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[-2i\pi(x_1+x_2)]\,\mathcal{Q}_p(0)\mathcal{G}_p(0), \tag{11}$$

where

$$\begin{aligned} \mathcal{Q}_p(0)^2 &= \frac{\Gamma(a+b-c+1)\Gamma(a-b+c+1)\Gamma(a+b+c+2)\Gamma(a+e-f+1)}{\Gamma(-a+b+c+1)}\,\frac{\Gamma(a-e+f+1)\Gamma(a+e+f+2)}{\Gamma(-a+e+f+1)}\\ &\times\frac{\Gamma(b+d-f+1)\Gamma(-b+d+f+1)\Gamma(b+d+f+2)}{\Gamma(b-d+f+1)}\,\frac{\Gamma(c+d-e+1)\Gamma(-c+d+e+1)\Gamma(c+d+e+2)}{\Gamma(c-d+e+1)}\\ &= \frac{\Gamma(\omega_{-})\Gamma(\omega_{01})\Gamma(\omega_{02})\Gamma(\omega^{13})\Gamma(\omega^{14})\Gamma(\omega^{15})\Gamma(\omega^{23})\Gamma(\omega^{24})\Gamma(\omega^{25})\Gamma(\omega^{34})\Gamma(\omega^{35})\Gamma(\omega^{45})}{\Gamma(\omega^{03})\Gamma(\omega^{04})\Gamma(\omega^{05})\Gamma(\omega_{12})} \end{aligned} \tag{12}$$

is the product of all the $\Gamma(\omega)$ limited to the ω's in which x_0 appears with a minus sign, the $\Gamma(\omega)$ of negativelike ω's being replaced by $\Gamma(1-\omega)^{-1}$.

B. Discussion

The classical symmetries of usual 6-j symbols are permutation of the columns of (9) and an even number of changes of signs among x_3, x_4, x_5. This group of 24 elements is extended to 144 by Regge's symmetries[11] which are equivalent to permutation of (x_3,x_4,x_5) independently of the one of (x_0,x_1,x_2). They are the only operations which do not mix negative like and positive like ω's.

TABLE I. The 17 patterns of parameters for the 160 ${}_4F_3(1)$ series, taking into account classical symmetries. There $\binom{ijk}{lmn}$ stands for $\binom{012}{345}$. Classical symmetries are permutations of columns and an even number of changes of signs for l,m,n. Patterns 3a and 3b are identical if Regge's symmetry (permutation of l,m,n independently of i,j,k) is taken into account. Idem for patterns 5,6,7, and 9. The last columns are respectively the multiplicity, the number of negativelike numerator parameters, how many of them appear in Shelepin's symbol, and the label of the related formulas in the text. (* = even number of minus signs, ** = odd number of minus signs)

1	$S(-i,-j,-k)$	1	4	0	35
2	$S(i,-j,-k)$	3	4	4	15
3a	$S(-i,-j,\pm l)$	12	4	2	33
3b	$S(-i,-j,\pm n)$	6	4	2	32
4	$S(l,\pm m,\pm n)$*	4	3	3	29
5a	$S(-i,\pm l,\pm m)$	24	3	2	30
5b	$S(-i,\pm m,\pm n)$	12	3	2	31
6a	$S(i,-j,\pm l)$	12	2	2	16
6b	$S(i,-j,\pm m)$	12	2	2	17
6c	$S(i,-j,\pm n)$	12	2	2	18
7a	$S(i,\pm l,\pm m)$	24	1	1	19
7b	$S(i,\pm m,\pm n)$	12	1	1	20
8	$S(\pm l,\pm m,\pm n)$**	4	1	0	34
9a	$S(i,j,\pm l)$	12	0	0	23
9b	$S(i,j,\pm n)$	6	0	0	24
10	$S(i,j,-k)$	3	0	0	22
11	$S(i,j,k)$	1	0	0	21

Consequently we shall obtain no further symmetry. These symmetries are similar to the one of the ${}_4F_3(1)$ series as discussed below (2), but cannot be obtained directly from an unique formula for the 6-j symbols because $S(0,1,2)$ and $S(-0,-1,-2)$ cannot be used to express them, as we shall see later. Other permutations and changes of sign among $(x_0 \ldots x_5)$ are related to Yutsis'mirror symmetry[10]: for example $x_0 \leftrightarrow x_3$ is $d \rightarrow -d-1$. Some of these symmetries have been derived[14] or discussed[15] from the properties of finite Saalschützian ${}_4F_3(1)$ series.

As can be seen in (11), the generalized 6-j symbol (7) is invariant for the permutation of x_1 and x_2, the permutations and an even number of changes of signs of (x_3,x_4,x_5). These symmetries are related to permutations of the columns and of the two first lines of Shelepin's symbol. Permutations with the third line define two other 6-j symbols related to $\mathcal{G}_p(1)$ and $\mathcal{G}_p(2)$. We shall extend to 12 6-j symbols using the notation

$$\overline{\mathcal{G}}_p(\lambda) = \mathcal{Q}_p(\lambda)\mathcal{G}_p(\lambda), \qquad \overline{\mathcal{G}}_n(\lambda) = \mathcal{Q}_n(\lambda)\mathcal{G}_n(\lambda), \qquad \mathcal{Q}_p(\lambda)\mathcal{Q}_n(\lambda) = 1, \tag{13}$$

where $\mathcal{Q}_p(\lambda)$ is defined as $\mathcal{Q}_p(0)$. Note that the numerator of $\mathcal{Q}_p(i)$ includes 12 Γ functions and its denominator only four of them; these figures are inverted for $\mathcal{Q}_n(i)$; the numerator and the denominator of $\mathcal{Q}_p(l)$ and $\mathcal{Q}_n(l)$ include eight Γ functions. The relation between the three $\overline{\mathcal{G}}_p(i)$ is simpler than the one between the three $\mathcal{G}_p(i)$ as follows,

$$\sin 2\pi(x_1-x_2)\overline{\mathcal{G}}_p(0) + \sin 2\pi(x_2-x_0)\overline{\mathcal{G}}_p(1) + \sin 2\pi(x_0-x_1)\overline{\mathcal{G}}_p(2) = 0. \tag{14}$$

We shall not list the other relations.

C. Formulas

We are primarily interested by a list of the finite ${}_4F_3(1)$ series which can be used to express a 6-j symbol. They are 160 $S(\pm\lambda,\pm\mu,\pm\nu)$ and we can already classify them with respect to the pattern of their parameters as shown on Table I. There are 11 patterns, taking into account Regge's symmetry, but 17 without it. Formulas of Ref. 7 are respectively of types 2, 6c, 3b, 6a and 3a.

Among the twenty expressions of $\overline{\mathcal{G}}_p(0)$ in terms of two ${}_4F_3(1)$ we can examine in which circumstances one of them reduces to a single term.

The patterns 2 and 11 of Table I appear in the expression with $S(0,1,2)$ and $S(0,-1,-2)$ which is the definition (7). $S(0,1,2)$ is always convergent but can never appear alone. On the contrary, if any one of $(a-b-c, a-e-f, -b+d-f, -c+d-e)$ is zero or a negative integer we obtain Rose's formula[13]:

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)]$$

$$\times \frac{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(b+c+e+f+2)}{\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(b-d+f+1)\Gamma(c-d+e+1)\Gamma(a-b+d-e+1)\Gamma(a-c+d-f+1)}$$

$$\times {}_4F_3[a-b-c,\ a-e-f,\ -b+d-f,\ -c+d-e;\ a-b+d-e+1,\ a-c+d-f+1,$$

$$-b-c-e-f-1;1]. \tag{15}$$

With all the parameters integer, the sum runs from $\max\{0,\ -a+b-d+e,\ -a+c-d+f\}$ to $\min\{-a+b+c,$ $-a+e+f, b-d+f, c-d+e\}$.

The pattern 6a is associated with the pattern 9a, using $S(0,1,-3)$ and $S(0,-1,3)$; if $a-b-c$ or $a-e-f$ is a negative integer, we obtain

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)] \frac{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(a+b+c+2)}{\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(b+d-f+1)\Gamma(b-d+f+1)}$$

$$\times \frac{\Gamma(a+e+f+2)\Gamma(-a+b+d+e+1)}{\Gamma(c-d+e+1)\Gamma(-c+d+e+1)\Gamma(2a+2)\Gamma(a-b+d-e+1)}$$

$$\times {}_4F_3[a-b-c, a-b+c+1, a-e-f, a-e+f+1; 2a+2, a-b-d-e, a-b+d-e+1; 1]. \quad (16)$$

With all the parameter integer, the sum is from $\max\{0, -a+b-d+e\}$ to $\min\{-a+b+c, -a+e+f\}$.

Using $S(0,1,-4)$ and $S(0,-1,4)$, if $a-e-f$ or $-c+d-e$ is a negative integer, we obtain the pattern 6b:

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)] \frac{\Delta(a,b,d)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)}{\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(a+e-f+1)\Gamma(b-d+f+1)}$$

$$\times \frac{\Gamma(a+b+c+2)\Gamma(b+d+f+2)\Gamma(2e+1)}{\Gamma(-c+d+e+1)\Gamma(c-d+e+1)\Gamma(a+b+d-e+2)\Gamma(a-b+d-e+1)}$$

$$\times {}_4F_3[a-e-f, a-e+f+1, -c+d-e, c+d-e+1; a+b+d-e+2, a-b+d-e+1, -2e; 1], \quad (17)$$

with, in the usual case, sum from $\max\{0, -a+b-d+e\}$ to $\min\{-a+e+f, c-d+e\}$. In neither of these two cases can the associated pattern 9a appear alone.

The pattern 6c is associated with the pattern 9b in the expression in terms of $S(0,1,-5)$ and $S(0,-1,5)$. If $a-e-f$ or $-b+d-f$ is zero or a negative integer, we obtain

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)]$$

$$\times \frac{\Delta(a,e,f)\Delta(b,d,f)\Gamma(a-b+c+1)\Gamma(c+d-e+1)\Gamma(b-c+e+f+1)}{\Delta(a,b,c)\Delta(c,d,e)\Gamma(-a+e+f+1)\Gamma(b-d+f+1)\Gamma(a-b+d-e+1)\Gamma(a+c+d-f+2)}$$

$$\times {}_4F_3[a-b+c+1, a-e-f, -b+d-f, c+d-e+1; -b+c-e-f, a-b+d-e+1,$$

$$a+c+d-f+2; 1], \quad (18)$$

with a summation from $\max\{0, -a+b-d+e\}$ to $\min\{-a+e+f, b-d+f\}$ when all parameters are integer. Formulas of type 6a, 6b, 6c given by (16)–(18) are equivalent by Regge's symmetry.

In the expression with $S(0,3,4)$ and $S(0,-3,-4)$, the coefficient of the first ${}_4F_3(1)$ series includes $\Gamma(-a-b+d+e)/\Gamma(-b+d-f)$ and the coefficient of the second one includes $\Gamma(a+b-d-e)/\Gamma(a-e-f)$. If $a-e-f$ is zero or a negative integer, the coefficient of the second ${}_4F_3(1)$ series vanishes and the ratio of Γ functions given above for the first ${}_4F_3(1)$ can be replaced by $(-)^{a-e-f}\Gamma(b-d+f+1)/\Gamma(a+b-d-e+1)$. We obtain a formula with the pattern 7a:

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(a+b+c)]$$

$$\times \frac{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(a+b+c+2)}{\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(-b+d+f+1)\Gamma(c-d+e+1)\Gamma(-c+d+e+1)}$$

$$\times \frac{\Gamma(a+e+f+2)\Gamma(b+d+f+2)}{\Gamma(2a+2)\Gamma(a+b+d-e+2)\Gamma(a+b-d-e+1)}$$

$$\times {}_4F_3[a+b-c+1, a+b+c+2, a-e-f, a-e+f+1; 2a+2, a+b+d-e+2,$$

$$a+b-d-e+1; 1], \quad (19)$$

with a summation between $\max\{0, -a-b+d+e\}$ and $-a+e+f$ in the usual case. If $-b+d-f$ is zero or a negative integer, instead of $a-e-f$ we obtain a formula which differs from (19) by the exchange of (a,b) with (d,e). If $a-e-f$ and $-b+d-f$ are both zero or negative integers, we can let them go to their value and get any one of these formulas. In Sec. 4 we shall get the same result by considering first $a+b-d-e$ as integer; in fact, formulas derived from $S(0,3,4)$ and $S(0,-3,-4)$ are identical term by term.

Similar considerations apply to the pattern 7b, obtained from $S(0,4,5)$:

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(a+b+c)]$$

$$\times \frac{\Delta(a,b,c)\Delta(a,e,f)\Gamma(a+b+c+2)\Gamma(b+d-f+1)\Gamma(c+d-e+1)}{\Delta(b,d,f)\Delta(c,d,e)\Gamma(-a+e+f+1)\Gamma(b+c-e-f+1)\Gamma(a+b+d-e+2)}$$

$$\times \frac{1}{\Gamma(a+c+d-f+2)} {}_4F_3[a+b+c+2,$$

$$a-e-f, b+d-f+1, c+d-e+1; a+b+d-e+2, a+c+d-f+2, b+c-e-f+1; 1], \tag{20}$$

with a summation from $\max\{0, -b-c+e+f\}$ to $-a+e+f$ in the usual case.

The ${}_4F_3(1)$ series which cannot be found alone are the one of pattern 11 which was introduced in the definition (7) and is

$$S(0,1,2) \sim {}_4F_3[a+b+c+2, a+e+f+2, b+d+f+2, c+d+e+2; a+b+d+e+3, a+c+d+f+3, b+c+e+f+3; 1], \tag{21}$$

the one of pattern 10 which is associated with itself,

$$S(0,1,-2) \sim {}_4F_3[a+b-c+1, a+e-f+1, b+d-f+1, -c+d+e+1; a+b+d+e+3, a-c+d-f+1, b-c+e-f+1; 1], \tag{22}$$

the one of pattern 9a,

$$S(0,1,3) \sim {}_4F_3[a+b+c+2, a+b-c+1, a+e+f+2, a+e-f+1; 2a+2, a+b+d+e+3, a+b-d+e+2; 1], \tag{23}$$

and 9b,

$$S(0,1,5) \sim {}_4F_3[a+b+c+2, a+e-f+1, b+d-f+1, c+d+e+2; a+b+d+e+3, a+c+d-f+2, b+c+e-f+2; 1]. \tag{24}$$

Formulas (15)–(20) give the generalized 6-j symbol as defined by (7) when they are finite series. The generalized 6-j symbol is invariant under permutation of the two last columns and the exchange of two lower quantum numbers with two upper ones; these symmetries are also valid to derive other formulas from (15)–(20).

4. EXISTENCE OF ONE INTEGER

There are relations between some $\bar{\mathcal{G}}_p(\lambda)$ and $\bar{\mathcal{G}}_n(\mu)$ when one of the ω's is zero or a negative integer, but also when some $2(x_\lambda \pm x_\mu)$ is an integer, as can be seen in (6). The ω's are numerator parameters of the ${}_4F_3(1)$ series and appear as the last five numerator parameters of the ${}_7F_6(1)$ series. The relations obtained when there are negative integers is the inversion of the finite series, which conserves the Saalschützian character of the ${}_4F_3(1)$ series and the special form of the well-poised ${}_7F_6$ series (the ${}_7F_6$ series is invariant in the inversion with respect to the first numerator parameter).

A. A denominator parameter is an integer

The denominator parameters are $1 \pm 2x_\lambda \pm 2x_\mu$. The relations obtained when they are an integer, q, come from the cancellation of the first few terms in the series in which they appear negative. These relations are:

$$\bar{\mathcal{G}}_p(i) = (-)^q \mathcal{G}_p(j), \quad \bar{\mathcal{G}}_n(i) = (-)^q \bar{\mathcal{G}}_n(j), \qquad \text{if } 2(x_i - x_j) = q, \tag{25a}$$

$$\bar{\mathcal{G}}_p(i) = \frac{\pi^4}{\sin\pi\omega_{\_}\sin\pi\omega_{kl}\sin\pi\omega_{km}\sin\pi\omega_{kn}} \bar{\mathcal{G}}_n(j), \qquad \text{if } 2(x_i + x_j) = q, \tag{25b}$$

$$\bar{\mathcal{G}}_p(i) = \frac{\pi^2}{\sin\pi\omega_{jl}\sin\pi\omega_{kl}} \bar{\mathcal{G}}_p(l), \quad \bar{\mathcal{G}}_n(l) = \frac{\pi^2}{\sin\pi\omega_{jl}\sin\pi\omega_{kl}} \bar{\mathcal{G}}_n(i), \qquad \text{if } 2(x_i - x_l) = q, \tag{25c}$$

$$\overline{\mathcal{G}}_p(i)=\frac{\pi^2}{\sin\pi\omega_{jm}\sin\pi\omega_{km}}\overline{\mathcal{G}}_n(l),\quad \overline{\mathcal{G}}_p(l)=\frac{\pi^2}{\sin\pi\omega_{jm}\sin\pi\omega_{km}}\overline{\mathcal{G}}_n(i),\quad \text{if } 2(x_i+x_l)=q, \tag{25d}$$

$$\overline{\mathcal{G}}_p(l)=\overline{\mathcal{G}}_p(m),\quad \overline{\mathcal{G}}_n(l)=\overline{\mathcal{G}}_n(m),\quad \text{if } 2(x_l-x_m)=q, \tag{25e}$$

$$\overline{\mathcal{G}}_p(l)=\overline{\mathcal{G}}_n(m),\quad \text{if } 2(x_l+x_m)=q. \tag{25f}$$

The difference between (25c) and (25d) is only a change of sign for x_l and x_m or x_n. These formulas are not unique because $\sin\pi\omega_{\lambda\mu}=\sin\pi\omega^{\lambda\mu}$; for example m can be replaced by n in (25d).

In the expression (5) of $\mathcal{G}_p(0)$ in terms of two ${}_4F_3(1)$ series, $\sin 2\pi(x_1+x_2)$ appears in the denominator; $2(x_1+x_2)$ is also the difference between the parameters of the two ${}_4F_3(1)$ series. Expanding them, we obtain derivatives when $2(x_1+x_2)$ goes to the integer value q:

$$\mathcal{G}_p(0)=\frac{1}{\prod_i\{\Gamma(a_i)\Gamma(a_i+q)\}}\left(\sum_{\nu=0}^{q-1}(-)^\nu\frac{(q-\nu-1)!}{\nu!}\frac{\prod_i\Gamma(a_i+\nu)}{\Gamma(b+\nu)\Gamma(c+\nu)}+(-)^q\right.$$

$$\times\sum_{\mu=0}^{\infty}\frac{\Pi_i\,\Gamma(a_i+q+\mu)}{(q+\mu)!\mu!\Gamma(b+q+\mu)\Gamma(c+q+\mu)}\Big[\psi(b+q+\mu)+\psi(c+q+\mu)+\psi(q+\mu+1)$$

$$\left.+\psi(\mu+1)-\sum_i\psi(a_i+q+\mu)\Big]\right), \tag{26}$$

where the a_i's are the numerator parameters of the first ${}_4F_3(1)$ series, b,c, and $1-q$ its denominator parameters, and ψ the logarithmic derivative of the Γ function. When one of the a_i's is a negative integer, $-p$, (26) reduces to a finite ${}_4F_3(1)$ series: if $p<q$, the sum of ν is limited by p; if $p>q$ the sum on ν disappears and the sum on μ reduces to an ordinary finite ${}_4F_3(1)$ series limited by $p-q$.

B. A numerator parameter is a negative integer

Any ω is a numerator parameter of twenty of the 160 ${}_4F_3(1)$ series involved. If one of them is a negative integer, relations between these twenty finite Saalschützian ${}_4F_3(1)$ series are known[16] and have already been used.[14,15] The same ω appears in ten expressions of six of the $\mathcal{G}_p$'s or $\mathcal{G}_n$'s in terms of two ${}_4F_3(1)$ series, and generates relations between these six functions.

If $\omega^*=-a-b-c-1$ is a negative integer,

$$\overline{\mathcal{G}}_n(l)=\frac{\pi^2\overline{\mathcal{G}}_n(i)}{\sin\pi\omega^{ij}\sin\pi\omega^{jk}}, \tag{27a}$$

where $l=3,4,5$ and $i=0,1,2$. If ω_{lm} is a negative integer ($\omega_{34}=-c-d-e-1$), we obtain the relation from (27a) by changing the signs of x_l and x_m,

$$\overline{\mathcal{G}}_p(l)=\overline{\mathcal{G}}_p(m)=\overline{\mathcal{G}}_n(n)=\frac{\pi^2\overline{\mathcal{G}}_n(i)}{\sin\pi\omega_{jn}\sin\pi\omega_{kn}}. \tag{27b}$$

If ω^{jk} is zero or a negative integer ($\omega^{12}=a-e-f$),

$$\overline{\mathcal{G}}_p(i)=\frac{\pi^2\overline{\mathcal{G}}_p(l)}{\sin\pi\omega_{jl}\sin\pi\omega_{kl}}=\frac{\pi^4\overline{\mathcal{G}}_n(j)}{\sin\pi\omega_{-}\sin\pi\omega_{kl}\sin\pi\omega_{km}\sin\pi\omega_{kn}}. \tag{27c}$$

If ω_{il} is zero or a negative integer ($\omega_{03}=a-b-c$), we obtain relation from (27c) by changing the signs of x_m and x_n,

$$\overline{\mathcal{G}}_p(i)=\frac{\pi^2\overline{\mathcal{G}}_p(l)}{\sin\pi\omega^{ij}\sin\pi\omega^{ik}}=\frac{\pi^2\overline{\mathcal{G}}_n(m)}{\sin\pi\omega_{jn}\sin\pi\omega_{kn}}=\frac{\pi^4\overline{\mathcal{G}}_n(j)}{\sin\pi\omega_{mn}\sin\pi\omega^{ij}\sin\pi\omega_{km}\sin\pi\omega_{kn}}. \tag{27d}$$

C. Formulas deduced from $\overline{G}_p(3)$

Formulas (15)–(20) give different expressions of the same 6-j symbol which hold when $a-e-f$ is zero or a negative integer. But, in this case

$$\overline{\mathcal{G}}_p(0)=\frac{\pi^2}{\sin\pi(c+d-e+1)\sin\pi(b+d-f+1)}\overline{\mathcal{G}}_p(3), \tag{28}$$

and $\overline{\mathcal{G}}_p(3)$ reduces to a single ${}_4F_3(1)$ series every time $a-e-f$ appears. The relation (28) is also the relation (25c) which holds when 2d is integer.

Writing $\overline{\mathcal{G}}_p(3)$ with $S(3,4,5)$, we obtain a pattern of type 4:

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(a+b+c)]$$

$$\times \frac{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(a+b+c+2)}{\Gamma(-a+e+f+1)\Gamma(-b+d+f+1)\Gamma(-c+d+e+1)\Gamma(a+b-d-e+1)}$$

$$\times \frac{1}{\Gamma(a+c-d-f+1)\Gamma(b+c-e-f+1)}$$

$$\times {}_4F_3[a+b+c+2, a-e-f, b-d-f, c-d-e;$$
$$a+b-d-e+1, a+c-d-f+1, b+c-e-f+1; 1], \tag{29}$$

with a summation from $\max\{0, -a-b+d+e, -a-c+d+f, -b-c+e+f\}$ to $\min\{-a+e+f, -b+d+f, -c+d+e\}$ in the usual case.

The pattern 5a is obtained with $S(3,-1,4)$ and gives

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(a+b+c)] \frac{\Delta(a,b,c)\Delta(a,e,f)\Gamma(a+b+c+2)\Gamma(b-d+f+1)\Gamma(c+d-e+1)}{\Delta(b,d,f)\Delta(c,d,e)\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(a+e-f+1)}$$

$$\times \frac{\Gamma(2e+1)\Gamma(-a+b+d+e+1)}{\Gamma(b+d+f+2)\Gamma(c+d+e+2)\Gamma(a+b-d-e+1)}$$

$$\times {}_4F_3[a-e-f, a-e+f+1, c-d-e, -c-d-e-1; a-b-d-e, a+b-d-e+1, -2e; 1], \tag{30}$$

with a sum from $\max\{0, -a-b+d+e\}$ to $\min\{-a+e+f, -c+d+e\}$ for SU2.

With $S(-2,3,4)$ we obtain pattern 5b which is

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp\{i\pi(a+b+c)\}$$

$$\times \frac{\Delta(a,e,f)\Delta(b,d,f)}{\Delta(a,b,c)\Delta(c,d,e)} \frac{\Gamma(a+b-c+1)\Gamma(-a+c+d+f+1)}{\Gamma(-a+e+f+1)\Gamma(-b+d+f+1)} \frac{\Gamma(-b+c+e+f+1)}{\Gamma(c+d+e+2)\Gamma(a+b-d-e+1)}$$

$$\times {}_4F_3[a+b-c+1, a-e-f, \quad b-d-f, -c-d-e-1;$$

$$a+b-d-e+1, a-c-d-f, \quad b-c-e-f; 1], \tag{31}$$

with a sum from $\max\{0, -a-b+d+e\}$ to $\min\{-a+e+f, -b+d+f\}$ when all the coefficients are integer.

The pattern 3b is obtained from $S(3,-1,-2)$,

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)]$$

$$\times \frac{\Delta(a,b,c)\Delta(a,e,f)\Gamma(-a+b+d+e+1)\Gamma(-a+c+d+f+1)}{\Delta(b,d,f)\Delta(c,d,e)\Gamma(-a+b+c+1)\Gamma(-a+e+f+1)\Gamma(b+d+f+2)} \frac{\Gamma(b+c+e+f+2)}{\Gamma(c+d+e+2)}$$

$$\times {}_4F_3[a-b-c, a-e-f, -b-d-f-1, -c-d-e-1; a-b-d-e, a-c-d-f,$$
$$-b-c-e-f-1; 1], \tag{32}$$

with a sum from 0 to $\min\{-a+b+c, -a+e+1\}$ when all the parameters are integer.

D. Other formulas

In order to obtain the pattern 3a when $a-e-f$ is a negative integer, we must use $\overline{\mathcal{G}}_p(4)$ which is also related to $\overline{\mathcal{G}}_p(0)$ in this case, and express it with $S(4,-1,-2)$:

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} = \exp[i\pi(b+c+e+f)] \frac{1}{\Gamma(-c+d+e+1)\Gamma(c-d+e+1)}$$

$$\times \frac{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(-b+c+e+f+1)\Gamma(b+c+e+f+2)\Gamma(2e+1)}{\Gamma(-a+b+c+1)\Gamma(a-b+c+1)\Gamma(-a+e+f+1)\Gamma(a+e-f+1)\Gamma(-b+d+f+1)\Gamma(b-d+f+1)}$$
$$\times {}_4F_3[a-e-f,\ -c+d-e,\ -a-e-f-1,\ -c-d-e-1;\ b-c-e-f,\ -b-c-e-f-1,\ -2e;\ 1], \tag{33}$$
with a summation from 0 to $\min\{-a+e+f, c-d+e\}$ in the usual case.

The pattern 8 can be obtained from $\mathcal{G}_p(3)$ expressed with $S(3,4,-5)$ and $S(3,-4,5)$ when $-c-d-e-1$ is also a negative integer

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix}$$
$$= -\exp[i\pi(b-d+f)]\frac{\pi^2}{\sin\pi(c+d-e+1)\sin\pi(b-d+f+1)}\frac{\Gamma(a+b-c+1)\Gamma(a-e+f+1)\Gamma(b-d+f+1)}{\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(c+d+e+2)}$$
$$\times\frac{1}{\Gamma(a+b-d-e+1)\Gamma(a-c-d+f+1)\Gamma(b-c-e+f+1)}{}_4F_3[a+b-c+1,\ a-e+f+1,\ b-d+f+1,$$
$$-c-d-e-1;\ a+b-d-e+1,\ a-c-d+f+1,\ b-c-e+f+1;\ 1]. \tag{34}$$

It cannot be used if $c+d+e$ or $b+d-f$ is zero or an integer, in which cases the ${}_4F_3(1)$ series vanishes.

To reach pattern 1, we must relate $\mathcal{G}_p(0)$ to one of the $\mathcal{G}_n(i)$. We obtain

$$\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix}$$
$$= -\exp[i\pi(a+b+c)]$$
$$\times\frac{\pi^2\sin\pi(a+b+d+e+1)}{\sin\pi(a+b-c+1)\sin\pi(a+e-f+1)}\frac{\Gamma(a+b+d+e+2)\Gamma(a+c+d+f+2)\Gamma(b+c+e+f+2)}{\sin\pi(b+d-f+1)\Delta(a,b,c)\Delta(a,e,f)\Delta(b,d,f)\Delta(c,d,e)\Gamma(a+b+c+2)}$$
$$\frac{1}{\Gamma(a+e+f+2)\Gamma(b+d+f+2)\Gamma(c+d+e+2)}{}_4F_3[-a-b-c-1,\ -a-e-f-1,\ -b-d-f-1,$$
$$-c-d-e-1;\ -a-b-d-e-1,\ -a-c-d-f-1,\ -b-c-e-f-1;\ 1], \tag{35}$$

valid if $a-e-f$ and one of $(-a-e-f-1,\ -b-d-f-1,\ -c-d-e-1)$ is a negative integer, but not if $-a-b-c-1$ is a negative integer, because this expression has been simplified by $\sin(a+b+c)$. It is clear that (34) and (35) cannot be used for SU2.

Formulas (29)–(33) are equivalent to (15)–(20) when $-a+e+f$ is an integer for any value of a,b,c,d,e,f. Their practical interest for SU2 is not the same. Combining the lower and upper limit of the summation, one can see that the total number of terms is one plus the smallest number of Shelepin's symbol in (15) and (29), one plus the lower of two of these numbers for (19), (20), and (32), one plus the lowest of four of these numbers for the others. The numerator parameters of the ${}_4F_3(1)$ series in (15) are on a line of Shelepin's symbols; those of (29) are related to a column as prolongated in (10). We have found (29) only once[17] in the literature. To a large extent the multiplicity of formulas is only apparent. When introduced in a ${}_4F_3(1)$ series, the cancellation of the lowest few terms generates another ${}_4F_3(1)$ series. In this sense there is only one formula (15) and one formula (29) related to (15) by the inversion of the order of terms; these terms are the ones of Racah's formula.[18] The cancellation of the lowest few terms does not change the pattern except for 6a and 6b, which are exchanged.

We shall not discuss the use of ${}_7F_6(1)$ series. Their last five numerator and denominator parameters are parameters of ${}_4F_3(1)$ series; the first ones are different, but we can expect no systematically improved formulas from them. For example, the limit of ${}_7F_6[a,1+\frac{1}{2}a,b,...;\frac{1}{2}a,1+a-b,...;1]$ is $\{2\times{}_6F_5[1,b,...;1-b,...;1]-1\}$ when a goes to zero.

5. EXISTENCE OF MORE THAN ONE INTEGER

A. General remarks

The relations (25) involve no contradiction if more than one of them are to be considered. Even all the $2(x_\lambda \pm x_\mu)$ are integers, whereas no ω is an integer, if $a,b,...,f$ are all half-integer; all the x's are half-integer and relations (25) become

$$(-)^{2(x_j-x_k)}\mathcal{G}_p(i)=\frac{\pi^2}{(\sin\pi\omega_-)^2}\{\mathcal{G}_p(l),\mathcal{G}_n(l)\}=(-)^{2(x_j-x_k)}\frac{\pi^4}{(\sin\pi\omega_-)^4}\mathcal{G}_n(i), \tag{36}$$

but there is no expression by a single ${}_4F_3(1)$ series.

If more than one ω is a negative integer, some of the relations (6) can break down. To study this, let us make the following remarks about the relations (27):

TABLE II. Compatibility for existence of more than one negative integer among the ω's. The multiplicity of each ω is given in parentheses.

negative integer	incompatible negative integer	compatible negative integer
ω^*	$\omega^{ij}(3)$	$\omega_{lm}(3),\omega_{il}(9)$
ω_{lm}	$\omega_{in}(3)$	$\omega^*(1),\omega_{ln}(2),\omega^{ij}(3),\omega_{il}(6)$
ω^{ij}	$\omega^*(1),\omega_{il}(6)$	$\omega_{lm}(3),\omega^{ik}(2),\omega_{kl}(3)$
ω_{il}	$\omega_{mn}(1),\omega^{ij}(2),\omega_{jm}(4)$	$\begin{cases}\omega^*(1),\omega_{lm}(2),\omega^{jk}(1),\\ \omega_{im}(2),\omega_{jl}(2)\end{cases}$

(a) They are relations between $\overline{\mathcal{G}}_n(\lambda)$ for λ of which the x_λ enters with a plus sign in the ω which generates this relation, and the $\overline{\mathcal{G}}_p(\mu)$ for μ, of which x_μ enters with a minus sign.

(b) The denominators include sinuses of all the negativelike ω which differ from the one which is an integer by four signs. There are three of them for the relations related to ω^* and ω_{lm}, seven for the other relations. If we restrict our considerations to the ω's which appear in the three lines of the regular Shelepin's symbol given by (10), the arguments of the sinus are those which are not in the same row or on the same column as the one which is an integer; if we add the last line, the rule is inverted for it. These values are given in Table II.

(c) A relation involves at most one $\overline{\mathcal{G}}_p(i)$, without denominator. With respect to $\overline{\mathcal{G}}_p(i)$, they are two sinuses in denominator for $\overline{\mathcal{G}}_p(l)$ and $\overline{\mathcal{G}}_n(l)$ and four sinuses for $\overline{\mathcal{G}}_n(i)$.

B. Two numerator parameters are integer

If ω and ω' are zero or negative integer and differ only by the signs of x_λ and x_μ, they can be numerator parameter of the same ${}_7F_6(1)$ or ${}_4F_3(1)$ series and they cannot be one in each ${}_4F_3(1)$ series in any expression of ${}_7F_6(1)$ series in terms of ${}_4F_3(1)$ series. Their difference $2(x_\lambda \pm x_\mu)$ is also integer. The relation between six $\overline{\mathcal{G}}_p$ or $\overline{\mathcal{G}}_n$ related to ω is extended to eight, including $\overline{\mathcal{G}}_p(\lambda)$, $\overline{\mathcal{G}}_n(\lambda)$, $\overline{\mathcal{G}}_p(\mu)$ and $\overline{\mathcal{G}}_n(\mu)$, in agreement with relation (25) related to $2(x_\lambda \pm x_\mu)$.

On the contrary, if only x_λ and x_μ have the same sign, positive for example, in the relations deduced from ω and the one deduced from ω', there are vanishing sinuses in the denominator of $\overline{\mathcal{G}}_n(\lambda)$ and $\overline{\mathcal{G}}_n(\mu)$ which are the only common families of these two relations. With $\omega_{14} = -a+b-c = -n$ and $\omega_{03} = a-b-c = -m$,

$$\overline{\mathcal{G}}_p(0) = \frac{\pi^2\overline{\mathcal{G}}_p(3)}{\sin\pi(-b+d+f+1)\sin\pi(-c+d+e+1)} = \frac{\pi^2\overline{\mathcal{G}}_n(4)}{\sin\pi(a+b-c+1)\sin\pi(a-e+f+1)}$$

$$= \frac{\pi^4\overline{\mathcal{G}}_n(1)}{\sin\pi(a+b-c+1)\sin\pi(a+e+f)\sin\pi(a+e-f+1)\sin\pi(-c+d+e+1)}$$

$$\simeq(\omega_{14})(-)^{a-b+c}\frac{\Gamma(0)\pi\overline{\mathcal{G}}_n(5)}{\sin\pi(a+e-f+1)}$$

$$\simeq(\omega_{14})(-)^{a-b+c}\frac{\Gamma(0)\pi^3\overline{\mathcal{G}}_n(2)}{\sin\pi(a+e+f)\sin\pi(a-e+f+1)\sin\pi(-b+d+f+1)}, \tag{37a}$$

$$\overline{\mathcal{G}}_p(1) = \frac{\pi^2\overline{\mathcal{G}}_n(3)}{\sin\pi(a+b-c+1)\sin\pi(b-d+f+1)} = \frac{\pi^2\overline{\mathcal{G}}_p(4)}{\sin\pi(-a+e+f+1)\sin\pi(-c+d+e+1)}$$

$$= \frac{\pi^4\overline{\mathcal{G}}_n(0)}{\sin\pi(a+b-c+1)\sin\pi(b+d+f)\sin\pi(b+d-f+1)\sin\pi(-c+d+e+1)}$$

$$\simeq(\omega_{03})(-)^{-a+b+c}\frac{\Gamma(0)\pi\overline{\mathcal{G}}_n(5)}{\sin\pi(b+d-f+1)}$$

$$\simeq(\omega_{03})(-)^{-a+b+c}\frac{\Gamma(0)\pi^3\overline{\mathcal{G}}_n(2)}{\sin\pi(-a+e+f+1)\sin\pi(b+d+f)\sin\pi(b-d+f+1)}. \tag{37b}$$

But $2c+1=-2(x_2+x_5)=n+m$ and (25d) gives

$$\overline{\mathcal{G}}_p(2)\simeq(-)^{a-b+c}\frac{\Gamma(0)\pi}{\sin\pi(c-d+e+1)}\overline{\mathcal{G}}_n(5), \tag{37c}$$

$$\overline{\mathcal{G}}_p(5)\simeq(-)^{a-b+c}\frac{\Gamma(0)\pi}{\sin\pi(c-d+e+1)}\overline{\mathcal{G}}_n(2). \tag{37d}$$

In all these relations $\overline{\mathcal{G}}_n(2)$ and $\overline{\mathcal{G}}_n(5)$ present a vanishing $\sin\pi(n+1)$ in the denominator, which can be replaced by $(-)^n\Gamma(0)/\pi$: In formulas (15)–(20) and (29)–(33) these sines have been replaced by Γ functions of which the ones with negative argument were eliminated with the ones arising for a specific formula. This rule of sign provides the same result after elimination of $\Gamma(0)$ when the negative argument is integer.

However, Whipple's results are no longer valid for $\overline{\mathcal{G}}_n(2)$ and $\overline{\mathcal{G}}_n(5)$, so we must consider the sixteen ${}_7F_6(1)$ series and the twenty expressions one by one in terms of ${}_4F_3(1)$ series. Six ${}_7F_6(1)$ series of each of these families do not include ω_{03} or ω_{14} as numerator parameters, of which two are convergent series which vanish leading to an undeterminated result. For the other ${}_7F_6(1)$ series, $\Gamma(0)$ can be eliminated with some other Γ functions; five of them, in each family, include ω_{03} as numerator parameter and verify (37a), the five others include ω_{14} and verify (37b). But, among the five ${}_7F_6(1)$ series with ω_{03} as numerator parameter, four include $-m-n$ as denominator parameter and these series can be restarted from the $(m+n+1)$th term to infinity; when these series converge [six of $\overline{\mathcal{G}}_n(5)$ and two for $\overline{\mathcal{G}}_n(2)$], they verify (37c) or (37d) to a sign. A finite series limited by ω_{03} plus its asymptotic part verifies (37b) instead of (37a). The relation (14) still holds between (37a), (37b), (37c). This behavior is related to whether ω_{03}, ω_{14}, or $\omega_{03}+\omega_{14}$ tends first to an integer. In the expressions in terms of two ${}_4F_3(1)$ series, ω_{03} and ω_{14} do not appear in six of them. In these cases, the ${}_4F_3(1)$ series cancel one another and the result is undetermined. In six other expressions ω_{03} and ω_{14} appear once in each ${}_4F_3(1)$ series: Each ${}_4F_3(1)$ series verifies relation (37a) and (37b), respectively, but their sums do not verify (37c) because their differences do it to a sign. In four cases ω_{03} appears in a ${}_4F_3(1)$ series without ω_{14} in the other: The finite ${}_4F_3(1)$ series verifies (37a) but, $-m-n$ being one of its denominator parameters, the sum beyond $m+n+1$ plus the finite and the infinite ${}_4F_3(1)$ series verify relation (37b). In (37a) and (37b) we noted by $\simeq(\omega_{14})$ the use of finite sum limited by ω_{14} or the use of finite sum limited by ω_{03} plus an infinite sum.

A similar situation holds when $\omega^{12}=a-e-f$ and $\omega^*=-a-b-c-1$ are negative integers. $\overline{\mathcal{G}}_n(1)$ and $\overline{\mathcal{G}}_n(2)$ are not uniquely defined. The description given above for $\overline{\mathcal{G}}_n(2)$ and $\overline{\mathcal{G}}_n(5)$ holds with the difference that all the infinite ${}_7F_6(1)$ series are divergent, and among the four series limited by ω^* with $\omega^*+\omega^{12}$ as denominator parameter three are convergent, but only one of the finite series limited by ω^{12} is.

C. More than two numerator parameters are integer

If three ω's are negative integer they must be on different lines of Shelepin's symbols and there must be at least two on the same column, or on different columns and at least two on the same line. In all cases, at least two of them are compatible; if there is incompatiblity with the third one, relations are similar to (37).

When the three integer ω's are on different lines and columns, the ω on the other line and other column is also an integer. For example, if $\omega_{03}=a-b-c$, $\omega_{14}=-a+b-c$, and $\omega_{25}=-a-b+c$ are integer, $\omega^*=-a-b-c-1$ is also integer. From ω_{25} we get

$$\begin{aligned}\overline{\mathcal{G}}_p(2)&=\frac{\pi^2\overline{\mathcal{G}}_n(5)}{\sin\pi(-b+d+f+1)\sin\pi(-a+e+f+1)}\\
&\simeq(\omega_{03})(-)^{-a+b+c}\frac{\Gamma(0)\pi\overline{\mathcal{G}}_n(4)}{\sin\pi(c+d-e+1)}\simeq(\omega_{14})(-)^{a-b+c}\frac{\Gamma(0)\pi\overline{\mathcal{G}}_n(3)}{\sin\pi(c-d+e+1)}\\
&\simeq(\omega_{03})(-)^{-a+b+c}\frac{\Gamma(0)\pi^3\overline{\mathcal{G}}_n(1)}{\sin\pi(c+d+e)\sin\pi(-a+e+f+1)\sin\pi(c-d+e+1)}\\
&\simeq(\omega_{14})(-)^{a-b+c}\frac{\Gamma(0)\pi^3\overline{\mathcal{G}}_n(0)}{\sin\pi(c+d+e)\sin\pi(-b+d+f+1)\sin\pi(c+d-e+1)},\end{aligned} \tag{38}$$

and similar relations for ω_{03} and ω_{14} which can be deduced from (37a) and (37b). No further relation can be deduced for ω^* because there is a vanishing denominator for all the $\overline{\mathcal{G}}_n$ in (38).

If three corners of a rectangle on Shelepin's symbol are integer, the fourth corner is also an integer. For example, if $\omega_{03}=a-b-c$, $\omega_{04}=-c+d-e$, $\omega_{13}=-c-d+e$ are integer, $\omega_{14}=-a+b-c$ is integer and

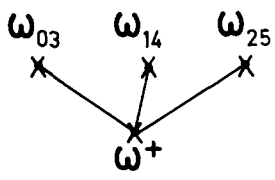

FIG. 1.

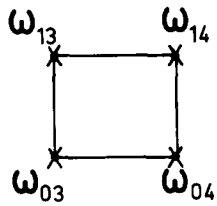

FIG. 2.

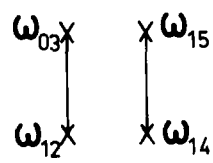

FIG. 3.

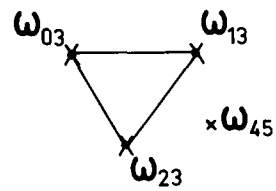

FIG. 4.

$$\overline{\mathcal{G}}_p(0) = (-)^{a+d-b-c}\overline{\mathcal{G}}_p(1) = \frac{\pi^2[\overline{\mathcal{G}}_p(3),\overline{\mathcal{G}}_p(4)]}{\sin\pi(-c+d+e+1)\sin\pi(-b+d+f+1)}$$

$$= \frac{\pi^2[\overline{\mathcal{G}}_n(3),\overline{\mathcal{G}}_n(4)]}{\sin\pi(a-e+f+1)\sin\pi(a+b-c+1)} = (-)^{a-b-c}\frac{\Gamma(0)\pi\overline{\mathcal{G}}_n(5)}{\sin\pi(b+d-f+1)}$$

$$= \frac{\pi^4[\overline{\mathcal{G}}_n(1),(-)^{a+d-b-e}\overline{\mathcal{G}}_n(0)]}{\sin\pi(a+e+f)\sin\pi(-c+d+e+1)\sin\pi(a+e-f+1)\sin\pi(a+b-c+1)}$$

$$= (-)^{a-b+c}\frac{\Gamma(0)\pi^3\overline{\mathcal{G}}_n(2)}{\sin\pi(a+e+f)\sin\pi(-b+d+f+1)\sin\pi(a-e+f+1)}. \tag{39}$$

There is no incompatibility: All the finite series lead to the result.

The existence of incompatibility can be deduced from some diagram displaying the integer ω's in a plane with lines between the compatible ones. The two cases discussed above are shown on Figs. 1 and 2. If this diagram is connex and has no free end, there is no incompatibility. If there are free ends or nonconnex parts some $\overline{\mathcal{G}}_p(\lambda)$ and $\overline{\mathcal{G}}_n(\lambda)$ can give different results, depending on which series is used. Further example with four integers are given in Figs. 3 and 4.

When all the parameters are integer, there is a relation between all the $\overline{\mathcal{G}}_p$'s and $\overline{\mathcal{G}}_n$'s which reads

$$(-)^{2(x_j-x_k)}\overline{\mathcal{G}}_p(i) = \Gamma(0)\Gamma(0)[\overline{\mathcal{G}}_p(l),\overline{\mathcal{G}}_n(l)] = (-)^{2(x_j-x_k)}\Gamma(0)\Gamma(0)\Gamma(0)\Gamma(0)\overline{\mathcal{G}}_n(i), \tag{40}$$

which relates all the finite sums.

D. Application to relations between Jacobi polynomials

In the hyperspherical formalism, let us consider the partition of Jacobi coordinates into three subsets of ν_i vectors ($i = 1,2,3$). The subset i is described by an hyperspherical polynomial of total quantum number K_i and a total length ρ_i but only $\alpha_i = K_i + \frac{3}{2}\nu_i - 1$ will appear in formulas. For a total degree $K = K_1 + K_2 + K_3 + 2\mathcal{N}$, we can couple the subset 1 to 2 with the quantum number n_{12} and then, (1,2) to 3 with the quantum number $N - n_{12}$, or alternatively 1 to 3 with n_{13} and (1,3) to 2 with $N - n_{13}$. Taking out the common factor, the transformation coefficients $\langle n_{13}|n_{12}\rangle_{\mathcal{N}}(\alpha_1,\alpha_2,\alpha_3)$ from one coupling to the other are defined by the relation

$$\eta_{n_{12}}^{\alpha_1,\alpha_3}\eta_{\mathcal{N}-n_{12}}^{\alpha_1+\alpha_2+2n_{12}+1,\alpha_3}\left(\frac{\rho_1^2+\rho_2^2}{\rho_1^2+\rho_2^2+\rho_3^2}\right)^{n_{12}}\mathcal{P}_{n_{12}}^{\alpha_1,\alpha_2}\left(\frac{\rho_2^2-\rho_1^2}{\rho_2^2+\rho_1^2}\right)\mathcal{P}_{\mathcal{N}-n_{12}}^{\alpha_1+\alpha_2+2n_{12}+1,\alpha_3}\left(\frac{\rho_3^2-\rho_2^2-\rho_1^2}{\rho_3^2+\rho_2^2+\rho_1^2}\right)$$

$$= \sum_{n_{13}=0}^{\mathcal{N}}\langle n_{13}|n_{12}\rangle_{\mathcal{N}}(\alpha_1,\alpha_2,\alpha_3)\eta_{n_{13}}^{\alpha_1,\alpha_2}\eta_{\mathcal{N}-n_{13}}^{\alpha_1+\alpha_3+2n_{13}+1,\alpha_2}\left(\frac{\rho_1^2+\rho_3^2}{\rho_1^2+\rho_2^2+\rho_3^2}\right)^{n_{13}}\mathcal{P}_{n_{13}}^{\alpha_1,\alpha_3}\left(\frac{\rho_3^2-\rho_1^2}{\rho_3^2+\rho_1^2}\right)\mathcal{P}_{\mathcal{N}-n_{13}}^{\alpha_1+\alpha_3+2n_{13}+1,\alpha_2}\left(\frac{\rho_2^2-\rho_3^2-\rho_1^2}{\rho_2^2+\rho_3^2+\rho_1^2}\right), \tag{41}$$

where $\mathcal{P}_n^{\alpha,\beta}(x)$ is a Jacobi polynomial and

$$\eta_n^{\alpha,\beta} = \left[\frac{2(2n+\alpha+\beta+1)n!\Gamma(n+\alpha+\beta+1)}{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}\right]^{1/2}, \tag{42}$$

a normalization constant. The transformation coefficient[5,6] is

$$\langle n_{13}|n_{12}\rangle_{\mathcal{N}}(\alpha_1,\alpha_2,\alpha_3) = \exp\{-i\pi(\alpha_1+\alpha_2+\alpha_3)\}[(2n_{12}+\alpha_1+\alpha_2+1)(2n_{13}+\alpha_1+\alpha_3+1)]^{1/2}$$

$$\times\begin{Bmatrix}\frac{1}{2}(\mathcal{N}+\alpha_1+\alpha_2) & n_{12}+\frac{1}{2}(\alpha_1+\alpha_2) & \frac{1}{2}\mathcal{N} \\ \frac{1}{2}(\mathcal{N}+\alpha_1+\alpha_3) & n_{13}+\frac{1}{2}(\alpha_1+\alpha_3) & \frac{1}{2}(\mathcal{N}+\alpha_2+\alpha_3)\end{Bmatrix}, \tag{43}$$

in the notation (7). The extended Shelepin's symbol (10) is

$$\begin{vmatrix} n_{13}+\alpha_1+\alpha_3 & n_{12}+\alpha_1 & n_{13}+\alpha_1 & n_{12}+\alpha_1+\alpha_2 \\ \mathcal{N}-n_{12}+\alpha_3 & \mathcal{N}-n_{13} & \mathcal{N}-n_{12} & \mathcal{N}-n_{13}+\alpha_2 \\ n_{13}+\alpha_3 & n_{12} & n_{13} & n_{12}+\alpha_2 \\ -\mathcal{N}-n_{12}-\alpha_1-\alpha_2-1 & -\mathcal{N}-n_{13}-\alpha_1-\alpha_2-\alpha_3-1 & -\mathcal{N}-n_{12}-\alpha_1-\alpha_2-\alpha_3-1 & -\mathcal{N}-n_{13}-\alpha_1-\alpha_3-1 \end{vmatrix}. \quad (44)$$

Relation (41) is valid between Jacobi polynomials for any values of $\alpha_1,\alpha_2,\alpha_3$. In this case, there are only four numerator parameters which are negative integer, and compatibility conditions are represented by Fig. 2. In the hyperspherical formalism, $\alpha_1,\alpha_2,\alpha_3$ are integer or half-integer, and one can use series limited by values other than n_{12}, n_{13}, $\mathcal{N}-n_{12}$ and $\mathcal{N}-n_{13}$, but some of them do not give the result when $\alpha_1,\alpha_2,\alpha_3$ are all half-integer.

It can easily be seen that $(-\mathcal{N}+n_{12},\ -\mathcal{N}+n_{13})$, $(-n_{12},\ -n_{13})$, $(-\mathcal{N}+n_{12},\ -n_{13})$ and $(-n_{12},\ -\mathcal{N}+n_{13})$ are common numerator parameters in four ${}_4F_3(1)$ series for each couple and $-n_{12}$, $-n_{13}$, $-\mathcal{N}+n_{12}$, $-\mathcal{N}+n_{13}$ appear alone in sixteen ${}_4F_3(1)$ series. Taking into account the trivial symmetry deduced from the permutation of 2 and 3, there are 44 formulas which look different, 12 with two negative numerator parameters and 32 with only one.

Moreover, as long as $\alpha_1,\alpha_2,\alpha_3$ are positive numbers, one of the four terms in the middle of Shelepin's symbol (44) is smaller than any other one. There are eight formulas (six, taking into account symmetry) of which the sum is limited by these four terms. For example, in (15) the sum runs from $\max\{0,\mathcal{N}-n_{12}-n_{13}\}$ to $\min\{n_{12},n_{13}\}$. Four of these formulas, $S(0,-1,-2)$, $S(-0,1,2)$, $S(3,-4,-5)$, and $S(-3,4,-5)$ involve twelve limitations for SU2; $S(0,-1,-5)$, $S(-0,1,-5)S(-2,3,-4)$ and $S(-2,-3,4)$ present only these four limitations. None of the two formulas of Ref. 6, which are $S(0,-2,-3)$ and $S(0,-2,-5)$ have this property.

6. SYMMETRIES

Formulas (15)–(20) and (29)–(33) can be used to obtain the 6-j symbol (7) only when the ${}_4F_3(1)$ series are finite; in general the 6-j symbols are related to some well-poised ${}_7F_6(1)$ series which corresponds to 120 of them by permutation of its arguments. There are twelve families of sixteen such ${}_7F_6(1)$ series with relations between three of them. So, there are symmetry relations between 23040 generalized 6-j symbols which are easily studied with the six x parameters (9).

(a) The definition (7) is invariant for a permutation of $(x_1 \ldots x_5)$ and Whipple's result is that it is invariant for an even number of changes of signs among $(x_1 \ldots x_5)$. But these symmetries do not respect the positivity of Shelepin's symbols or the usual SU2 relations between the arguments of the 6-j symbol (the argument of some Γ functions in the square root $\mathcal{Q}_p$ becomes negative). With respect to these requirements, the x's can be divided into two subsets such that $(x_0,x_1,x_2)<(x_3,x_4,x_5)$ and we can change only the sign of (x_3,x_4,x_5). The change of two signs among (x_3,x_4,x_5) is the permutation of two upper parameters with the corresponding lower ones in the 6-j symbol. Permutation (x_1,x_2) and (x_4,x_5) is the permutation of the two last columns. Other permutations of (x_3,x_4,x_5) are Regge's symmetries. These symmetries allowed with respect to SU2 correspond to permutations of columns and of the two first lines of Shelepin's symbol (10). They hold for any value of $a \cdots f$.

(b) x_0 can be permuted with x_1 if the difference is half-integer:

$$\exp[-2i\pi(x_1+x_2)]\,\overline{\mathcal{G}}_p(0) \to \exp[-2i\pi(x_0+x_2)]\,\overline{\mathcal{G}}_p(1)$$
$$= \exp[-2i\pi(x_0+x_2)]\left(\frac{\sin 2\pi(x_1-x_2)}{\sin 2\pi(x_0-x_2)}\overline{\mathcal{G}}_p(0)+\frac{\sin 2\pi(x_0-x_1)}{\sin 2\pi(x_0-x_2)}\overline{\mathcal{G}}_p(2)\right), \quad (45)$$

using relation (14). So the condition for symmetry in a permutation of (x_0,x_1,x_3), is $a+d-b-e$ and $a+d-c-f$ integers. With these two conditions all the symmetries symbolised by Shelepin's symbol are allowed.

(c) Other permutations and changes of sign among the x's are related to Yutsis' "mirror symmetry"[10] or "analytical continuation." However, the changes $j \to -j-1$ on the 144 "usual" 6-j symbols generate only a total of 9216 elements, that is, two out of five for the total number. Note that the same ratio holds for 3-j symbols. The exchange of x_2 and x_5 is $f \to -f-1$; $\overline{\mathcal{G}}_p(0)$ being invariant, we get

$$\begin{Bmatrix} a & b & c \\ d & e & -f-1 \end{Bmatrix} = \exp[i\pi(-2f-1)]\left(\frac{\sin\pi(a-e+f)\sin\pi(-b+d+f)}{\sin\pi(a-e-f)\sin\pi(-b+d-f)}\right)^{1/2}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix}, \quad (46)$$

to a sign for the square root. Considerations of Ref. 1 can be used here. The two 6-j symbols of (46) are equal to a sign for any integer or half integer value of f. For $f=-\frac{1}{2}$ they are identical, and we can follow the sign of the square root in an analytic continuation of f from $-\frac{1}{2}$ to its actual value and get the phase for a path which is symmetric for $f \to -f-1$ in the complex plane. The result is

$$\begin{Bmatrix} a & b & c \\ d & e & -f-1 \end{Bmatrix} = (-)^{(f+\frac{1}{2})\{\mathrm{Sign}[\mathrm{Im}(a-e)]-\mathrm{Sign}[\mathrm{Im}(-b+d)]\}}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix}, \quad (47)$$

for any value of $a\cdots e$ when f is integer or half-integer. The path can be deformed; then Sign[Im$(a-e)$] = 1 means that the zeros $-a+e+n\pi$ are above the path and the poles $a-e+n\pi$ below it. In the interchange of (x_1,x_2) with (x_4,x_5) $e\to -e-1$ and $f\to -f-1$; the phase is the product of the phase for $f\to -f-1$ with the phase of $e\to -e-1$ with f changed into $-f-1$. As f and e must be half integers, we can define the phase with respect to the imaginary parts of a, $-b+d$ and $c-d$; roles of e and f can be interchanged. In this particular case Yutsis[10] gets a complex phase. The exchange of x_0 and x_3 is $d\to -d-1$; a formula similar to (47) is obtained when d is integer or half-integer: $\overline{\mathcal{G}}_p(0)$ can be expressed in terms of $\overline{\mathcal{G}}_p(1)$ and $\overline{\mathcal{G}}_p(2)$ by (14); after the change $d\to -d-1$ in $\overline{\mathcal{G}}_p(1)$ and $\overline{\mathcal{G}}_p(2)$, (14) can be used back only for half-integer and integer values of d.

(d) It becomes very difficult to define the phase for the mirror symmetry of many arguments of the 6-j symbol by analytic continuation. Conversely some "mirror symmetries" extend to any value of the arguments if there are integer relations between other arguments of the 6-j symbol. For example, when $a-e$ and $b-d$ are integers, (46) becomes

$$\begin{Bmatrix} a & b & c \\ d & e & -f-1 \end{Bmatrix} = \exp[i\pi(-2f-1)] \begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix}, \tag{48}$$

for any value of f. Discussion of analytic continuation when $a-e$ and $b-d$ are integers or half-integers is similar to the one given for 3-j symbols.[1]

7. RECURRENCE RELATIONS

Recurrence relations are easily obtained between contiguous generalized hypergeometric series. Writing one of these relations for the well-poised ${}_7F_6(1)$ series as the one in (5) in terms of Whipple's parameters, we get relations between ${}_7F_6(1)$ series which are not contiguous. In fact, we go from the notion of contiguous series to the notion of contiguity in the x's parameter space.

The well-poised ${}_7F_6(1)$ series involved are

$$F = {}_7F_6[a,1+\tfrac{1}{2}a,b_1,\ldots,b_5;\tfrac{1}{2}a,1+a-b_1,\ldots,1+a-b_5;1], \tag{49}$$

and we are interested by a relation between F and two F^i where F^i is F in which b_i is replaced by b_i+1. Noting $x(F)$ and $x(F_i)$, Whipple's parameters of F and F_i,

$$x_j(F^i) = x_j(F) + \tfrac{1}{4} - \tfrac{1}{2}\delta_{ij}, \tag{50}$$

that is, all x's are increased by $\frac{1}{4}$ except for x_i, which decreases by $\frac{1}{4}$. The relation

$$(x_i-x_j)\omega_{ij}F + (x_0-x_i)\omega_{0i}F^i + (x_j-x_0)\omega_{0j}F^j = 0 \tag{51}$$

can be verified for each term of the series. Noting by $\mathcal{G}_p^{(i)}(0)$ in function $\mathcal{G}_p(0)$ defined by (5) in which all the x's are increased by $\frac{1}{4}$ except for x_i which is decreased by $\frac{1}{4}$,

$$2(x_i-x_j)\omega^*\omega_{ij}\mathcal{G}_p(0) + \omega_{0i}\mathcal{G}_p^{(i)}(0) - \omega_{0j}\mathcal{G}_p^{(j)}(0) = 0. \tag{52}$$

In this last relation, permutations of $(x_1\cdots x_5)$ and an even number of changes of signs are allowed. A relation between $\mathcal{G}_p^{(0)}(0)$ and $\mathcal{G}_p^{(i)}(0)$ can be obtained, expressing the $\mathcal{G}_p(0)$'s in terms of $\mathcal{G}_p(1)$ and $\mathcal{G}_p(2)$ and going back to $\mathcal{G}_p(0)$,

$$2(x_i-x_0)\omega^*\omega_{0i}\mathcal{G}_p(0) + \omega_{0i}\mathcal{G}_p^{(i)}(0) - \left\{\prod_{j=1}^{5}\omega_{0j}\right\}\mathcal{G}_p^{(0)}(0) = 0. \tag{53}$$

Combining (52) and (53) with those obtained with an even number of change of signs, we obtain a relation between $\mathcal{G}_p(0)$ and any of its 32 neighbors in the x's space; these neighbors are defined by the decrease of an odd number of x's by $\frac{1}{4}$ and the increase of the other by $\frac{1}{4}$. In all of these relations, the coefficient of $\mathcal{G}_p(0)$ separates in two parts such that

$$K^{(i,j\cdots)}\mathcal{G}_p(0) + M^{(i,j\cdots)}\mathcal{G}_p^{(i,j\cdots)}(0) = C(x), \tag{54}$$

where

$$K^{(i,j\cdots)} = 2y\omega^*(\omega^*-2y) + \tfrac{8}{3}(y^3-Y), \quad y=\sum_\lambda x_\lambda, \quad Y=\sum_\lambda x_\lambda^3, \quad \lambda\in i,j,\cdots, \tag{55}$$

and

$$M^{(i,j\cdots)} = \prod\omega_{0\lambda\mu\cdots} \tag{56}$$

is the product of all $\omega_{0\lambda\mu}$ such that $(0,\lambda,\mu,\cdots)$ differs from $(i,j,\cdots)$ by only one element; there is only one such ω if 0 is not in $(i,j,\cdots)$ and five such if 0 is. This relation has been obtained by looking at a tedious list of all possible cases. It is easier to show that it always holds: Since (52) and (53) agree with (54) and

$$K^{(i,j,\cdots B)}(x_\lambda=-a, x_\mu=-b) - K^{(i,j\cdots)}(x_\lambda=a, x_\mu=b) = 2(a+b)(\omega^*-2a)(\omega^*-2b) \tag{57}$$

depends only on the x's of which the sign has been changed, all relations deduced from (52) and (53) by an even number of

TABLE III. Relations between variations of x_0, x_3 and variations of a and d.

x_0	x_3	a	d	x_0	x_3	a	d
$x_0+\frac{1}{4}$	$x_3+\frac{1}{4}$	$a-\frac{1}{2}$	d	$x_0-\frac{1}{4}$	$x_3+\frac{1}{4}$	a	$d+\frac{1}{2}$
$x_0+\frac{1}{4}$	$x_3-\frac{1}{4}$	a	$d-\frac{1}{2}$	$x_0-\frac{1}{4}$	$x_3-\frac{1}{4}$	$a+\frac{1}{2}$	d

changes of sign among the x's and all relations deduced by their combinations agree with (54). $C(x)$ is a function of the x's which is not invariant when the sign of some of them is changed.

Introducing the square roots, the relation becomes for $\overline{\mathcal{G}}_p(0)$,

$$K^{(i,j,\cdots)}\overline{\mathcal{G}}_p(0) + N^{(i,j,\cdots)}\overline{\mathcal{G}}_p{}^{(i,j,\cdots)}(0) = \overline{C}(x) \tag{58}$$

where $K^{(i,j,\cdots)}$ is given by (55) and

$$N^{(i,j,\cdots)} = (-)^{\eta(i,j,\cdots)}\left\{\prod_1^6 |\omega^{\lambda\mu\cdots}|\right\}^{1/2}, \tag{59}$$

where $\{\lambda,\mu,\cdots\}$ differs from $(i,j,\cdots)$ by only one term; there are always six ω's in the square root. There is a change of sign for negativelike ω's noted by $|\omega^{\lambda\mu\cdots}|$. Here $\eta(i,j,\cdots)$ is the number of negativelike ω's in (56). Between generalized 6-j symbols the sign is different, due to the phase $\exp\{-2i\pi(x_1+x_2)\}$: For them the sign is plus if there are zero or one (0,1,2) among $(i,j\cdots)$ and minus if there are two or three of them.

Another criterium for the choice of the ω's in the square root (59) which involves no change of sign is to consider the positive ω's related to $\overline{\mathcal{G}}_p(0)$ and $\overline{\mathcal{G}}_p{}^{(i,j,\cdots)}(0)$; for six of them, the real part varies by 1: the one with the smallest real part are in the square root.

Variations of the arguments of the same column of a generalized 6-j symbol are related only to variations of the two related x's as shown on Table III. All recurrence relations obtained can be written

$$-(b+c-d)(a+d+1)(a+b+c+1)\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} + \{(a+b+c+1)(-a+b+c)(b-d+f)(-b+d+f+1)$$

$$\times(c-d+e)(-c+d+e+1)\}^{1/2}\begin{Bmatrix} a & b-1/2 & c-1/2 \\ d+1/2 & e & f \end{Bmatrix}$$

$$= -(b+c+d+1)(a-d)(a+b+c+1)\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} + \{(a+b+c+1)(-a+b+c)(b+d+f+1)(b+d-f)$$

$$\times(c+d+e+1)(c+d-e)\}^{1/2}\begin{Bmatrix} a & b-1/2 & c-1/2 \\ d-1/2 & e & f \end{Bmatrix}$$

$$= \{(a+b+c+d+e+f+3)(d+e+f+2)(a+b+c+1) - (a+b+d+e+2)(a+c+d+f+2)(b+c+e+f+2)\}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} - \{(a+e+f+2)(-a+e+f+1)(b+d+f+2)(-b+d+f+1)(c+d+e+2)$$

$$\times(-c+d+e+1)\}^{1/2}\begin{Bmatrix} a & b & c \\ d+1/2 & e+1/2 & f+1/2 \end{Bmatrix}.$$

$$= \{(2a+b+e+2)(a-c+e+1)(a+b+c+1) - (2a+1)(a+b+d+e+2)(a+b-d+e+1)\}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix}.$$

$$-\{(a+b-c+1)(-a+b+c)(a+e+f+2)(a+e-f+1)(-c+d+e+1)(c+d-e)\}^{1/2}$$

$$\times\begin{Bmatrix} a+1/2 & b & c-1/2 \\ d & e+1/2 & f \end{Bmatrix}$$

$$= \{(a+b+c+d+e-f+2)(d+e-f+1)(a+b+c+1) - (a+b+d+e+2)(a+c+d-f+1)(b+c+e-f$$

$$+1)\}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} - \{(a+e-f+1)(a-e+f)(b+d-f+1)(b-d+f)(c+d+e+2)(-c+d+e+1)\}^{1/2}$$

$$\times \begin{Bmatrix} a & b & c \\ d+1/2 & e+1/2 & f-1/2 \end{Bmatrix}$$

$$= \{(2a+b-e+1)(a-c-e)(a+b+c+1) - (2a+1)(a+b+d-e+1)(a+b-d-e)\}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} + \{(a+b$$

$$-c+1)(-a+b+c)(a-e+f+1)(-a+e+f)(c+d+e+1)(c-d+e)\}^{1/2}\begin{Bmatrix} a+1/2 & b & c-1/2 \\ d & e-1/2 & f \end{Bmatrix}$$

$$= \{(a+b+c+d-e-f+1)(d-e-f)(a+b+c+1) - (a+b+d-e+1)(a+c+d-f+1)(b+c-e-f)\}$$

$$\times\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} + \{(a+e+f+1)(-a+e+f)(b+d-f+1)(b-d+f)(c+d-e+1)(c-d+e)\}^{1/2}$$

$$\times \begin{Bmatrix} a & b & c \\ d+1/2 & e-1/2 & f-1/2 \end{Bmatrix}$$

$$= \{(-a-b-c+d+e+f)(d+e+f+1)(a+b+c+1) - (a+b-d-e)(a+c-d-f)(b+c-e-f)\}$$

$$\times\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} + \{(a+e+f+1)(-a+e+f)(b+d+f+1)(-b+d+f)(c+d+e+1)(-c+d+e)\}^{1/2}$$

$$\times \begin{Bmatrix} a & b & c \\ d-1/2 & e-1/2 & f-1/2 \end{Bmatrix}$$

$$= \{(2a+b+c-d+1)(a+2b+2c+d+3)(a+b+c+2) - T\}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} - \{(a+b+c+2)(-a+b+c$$

$$+1)(b+d+f+2)(b+d-f+1)(c+d+e+2)(c+d-e+1)\}^{1/2}\begin{Bmatrix} a & b+1/2 & c+1/2 \\ d+1/2 & e & f \end{Bmatrix}$$

$$= \{(2a+b+c-d+2)(a+2b+2c+d+3)(a+b+c+2) - T\}\begin{Bmatrix} a & b & c \\ d & e & f \end{Bmatrix} - \{(a+b+c+2)(-a+b+c$$

$$+1)(b-d+f+1)(-b+d+f)(c-d+e+1)(-c+d+e)\}^{1/2}\begin{Bmatrix} a & b+1/2 & c+1/2 \\ d-1/2 & e & f \end{Bmatrix},$$

with

$$T = 8(a+b+1)(a+c+1)(b+c+1) + (2a+1)(a+d+1)(a-d) + (2b+1)(b+e+1)(b-e) + (2c+1)(c+f+1)(c-f). \tag{60}$$

The relations (60) involve $\mathcal{G}_p^{(0)}, \mathcal{G}_p^{(3)}, \mathcal{G}_p^{(0,1,2)}, \mathcal{G}_p^{(0,1,3)}, \mathcal{G}_p^{(0,1,5)}, \mathcal{G}_p^{(0,3,4)}, \mathcal{G}_p^{(0,4,5)}, \mathcal{G}_p^{(3,4,5)}, \mathcal{G}_p^{(0,1,2,4,5)}$, and $\mathcal{G}_p^{(1,2,3,4,5)}$. They are completed by permutations of couples (ad), (be) and (cf). They hold for any complex value of $a,b,\ldots,f$ with the convention of positive square root at the real limit. The coefficient of the central 6-j symbol is simple when only one of the x's decreases, more complicated when three of them decrease, and involves T when only one increases. With a change of sign for an even number of the x's, the coefficient of the central 6-j symbols changes according to (57). However, for a relation written between any three 6-j symbols, the coefficient of the central term is invariant and can be factorized in some cases.

Three 6-j symbols with the same arguments, except for one which is respectively $a-1$, a and $a+1$, are not contiguous. The recurrence relation between them is not one of those studied here, but can be obtained by combining them.

8. SUMMARY AND CONCLUSION

A generalized 6-j symbol is defined by two sets of three parameters (x_0,x_1,x_2) and (x_3,x_4,x_5) where x_0 plays a special role. To compute each of them, there are 16 ${}_7F_6(1)$ series related to an even number of change of signs among $(x_1\cdots x_5)$ and there are 12 independent definitions with respect to what is selected to be x_0 and its sign. There are three term relations between any three of these independent definitions. However, if there are some integer relations between the x's, the 12 definitions become identical and the 6-j symbol can be expressed as the Saalschützian ${}_4F_3(1)$ series, among which there are 11 different patterns for the usual conditions of angular momenta.

Among the permutation of $(x_1\cdots x_5)$ and even number of change of signs, only permutation of (x_1,x_2) and permutations and changes of two signs among (x_3,x_4,x_5) preserve the usual triangular relations among the arguments of the generalized 6-j symbol. When $x_0 - x_1$ or $x_0 - x_2$ are half-integer the symmetry is extended to the permutations of (x_0,x_1,x_2) and involves the 144 elements of usual and Regge's symmetry.

Keeping x_0 fixed, permutation of the other x's give relation to other generalized 6-j symbols which are usually considered as analytic continuations of the usual ones. When the permuted x's differ by half an integer, these relations reduce to a plus or a minus sign according to the path of analytic continuation.

For many problems in which a coefficient is obtained as a finite Saalschützian ${}_4F_3(1)$ series, this study can be used to relate different formulas. However, we had to deal with a square root of which the structure is related to the existence of a Shelepin's symbol including positive values. For a given set of x's there is not necessarily a generalized 6-j symbol.

APPENDIX: RECURRENCE RELATIONS FOR GENERALIZED 3-j SYMBOLS

The special form of recurrence relations obtained here for generalized 6-j symbol incites us to look for a similar form for the recurrence relations of generalized 3-j symbols given in Appendix B of Ref. 1. The generalized 3-j symbols were written

$$\begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} = \exp[i\pi(r_5 - r_4)]\bar{F}_p(0), \tag{A1}$$

with

$$r_1 = \tfrac{1}{6}(3 + 6a + 2\gamma - 2\beta), \qquad r_2 = \tfrac{1}{6}(3 + 6b + 2\alpha - 2\gamma), \qquad r_3 = \tfrac{1}{6}(3 + 6c + 2\beta - 2\alpha),$$
$$r_4 = \tfrac{1}{6}(-3 - 6a + 2\gamma - 2\beta), \quad r_5 = \tfrac{1}{6}(-3 - 6b + 2\alpha - 2\gamma), \quad r_0 = \tfrac{1}{6}(-3 - 6c + 2\beta - 2\alpha). \tag{A2}$$

A contiguous 3-j symbol is obtained for four shifts of $\frac{1}{3}$ and two of $-\frac{2}{3}$ on r's parameters, or four shifts of $-\frac{1}{3}$ and two shifts of $\frac{2}{3}$. They were denoted by $\bar{F}_p^{(i,j)}(0)$ when r_i and r_j increase by $\frac{2}{3}$ and by $\bar{F}_p^{-(i,j)}(0)$ when r_i and r_j decrease by $\frac{2}{3}$. The recurrence obtained in Ref. 1 can be written

$$K^{i,j}(r)\bar{F}_p(0) - \epsilon(i,j)M^{i,j}(r)\bar{F}_p^{(i,j)}(0) = -K^{i,j}(-r)\bar{F}_p(0) + \epsilon(i,j)M^{i,j}(-r)\bar{F}_p^{-(i,j)}(0) = C(r),$$

with

$$K^{i,j}(r) = \tfrac{1}{2}(r_i + r_j + \tfrac{1}{2})^2 + \tfrac{1}{2}(r_i^2 + r_j^2) - \tfrac{1}{4}\sum_{\lambda=0}^{5} r_\lambda^2, \tag{A4}$$

$$M^{i,j}(r) = \{\prod_{\mu\neq i,j}(\tfrac{1}{2} + r_i + r_j + r_\mu)\}^{1/2}, \tag{A5}$$

$$\epsilon(i,j) = -1 \quad \text{if} \quad (i,j)\in(1,2,3), \qquad \epsilon(i,j) = 1 \quad \text{otherwise.} \tag{A6}$$

Here $K(-r)$, $M(-r)$ means that there must by a change of sign for all r's in (A4) or (A5).

When written for generalized 3-j symbols, the phase (A6) is $\epsilon(i,j) = -1$ if $(i,j)\in(0,4,5)$ and $\epsilon(i,j) = 1$, otherwise (there is an error on this point in Ref. 1). The 30 recurrence relations can be obtained by circular permutation of (a,b,c) in the ten following ones:

$$\{\tfrac{1}{2}(a + b - c + 1)(a - b - c) + (a + \tfrac{2}{3})\gamma + (c + \tfrac{1}{3})\alpha\}$$
$$\times\begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} - \{(a + b - c + 1)(-a + b + c)(a - \alpha + 1)(c - \gamma)\}^{1/2}\begin{pmatrix} a + 1/2 & b & c - \tfrac{1}{2} \\ \alpha - \tfrac{1}{2} & \beta & \gamma + \tfrac{1}{2} \end{pmatrix}$$

$$= \{\tfrac{1}{2}(a + b - c + 1)(-a + b - c) - (b + \tfrac{2}{3})\gamma - (c + \tfrac{1}{3})\beta\}\begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} - \{(a + b - c + 1)(a - b + c)(b + \beta + 1)$$
$$\times(c + \gamma)\}^{1/2}\begin{pmatrix} a & b + \tfrac{1}{2} & c - \tfrac{1}{2} \\ \alpha & \beta + \tfrac{1}{2} & \gamma - \tfrac{1}{2} \end{pmatrix}$$

$$= \{ -\tfrac{1}{2}[a(a+1)+b(b+1)-c(c+1)] - \alpha\beta - \tfrac{1}{3}(\alpha-\beta)\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} - \{(a+\alpha)(a-\alpha+1)(b-\beta)(b+\beta+1)\}^{1/2}$$
$$\times \begin{pmatrix} a & b & c \\ \alpha-1 & \beta+1 & \gamma \end{pmatrix}$$
$$= \{\tfrac{1}{2}(a+b+c+1)(a-b+c) - (a+\tfrac{1}{3})\gamma + (c+\tfrac{1}{3})\alpha\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} + \{(a+b+c+1)(a-b+c)(a+\alpha)(c-\gamma)\}^{1/2}$$
$$\times \begin{pmatrix} a-\frac{1}{2} & b & c-\frac{1}{2} \\ \alpha-\frac{1}{2} & \beta & \gamma+\frac{1}{2} \end{pmatrix}$$
$$= \{\tfrac{1}{2}(a+b+c+2)(a+b-c+1) - (a+\tfrac{2}{3})\beta + (b+\tfrac{2}{3})\alpha\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix}$$
$$- \{(a+b+c+2)(a+b-c+1)(a+\alpha+1)(b-\beta+1)\}^{1/2} \begin{pmatrix} a+\frac{1}{2} & b+\frac{1}{2} & c \\ \alpha+\frac{1}{2} & \beta-\frac{1}{2} & \gamma \end{pmatrix}$$
$$= \{\tfrac{1}{2}(-a+b+c+1)(a+b-c) - (a+\tfrac{1}{3})\gamma - (c+\tfrac{2}{3})\alpha\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix}$$
$$+ \{(-a+b+c+1)(a+b-c)(a-\alpha)(c-\gamma+1)\}^{1/2}$$
$$\times \begin{pmatrix} a-\frac{1}{2} & b & c+\frac{1}{2} \\ \alpha+\frac{1}{2} & \beta & \gamma-\frac{1}{2} \end{pmatrix}$$
$$= \{\tfrac{1}{2}(a-b+c+1)(a+b-c) + (b+\tfrac{1}{3})\gamma + (c+\tfrac{2}{3})\beta\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix}$$
$$+ \{(a-b+c+1)(a+b-c)(b+\beta)(c+\gamma+1)\}^{1/2} \begin{pmatrix} a & b-\frac{1}{2} & c+\frac{1}{2} \\ \alpha & \beta-\frac{1}{2} & \gamma+\frac{1}{2} \end{pmatrix}$$
$$= \{\tfrac{1}{2}[a(a+1)+b(b+1)-c(c+1)] + \alpha\beta - \tfrac{1}{3}(\alpha-\beta)\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix} + \{(a-\alpha)(a+\alpha+1)(b+\beta)(b-\beta+1)\}^{1/2}$$
$$\times \begin{pmatrix} a & b & c \\ \alpha+1 & \beta-1 & \gamma \end{pmatrix}$$
$$= \{\tfrac{1}{2}(a+b+c+2)(-a+b-c-1) + (a+\tfrac{2}{3})\gamma - (c+\tfrac{2}{3})\alpha\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix}$$
$$- \{(a+b+c+2)(a-b+c+1)(a+\alpha+1)(c-\gamma+1)\}^{1/2} \begin{pmatrix} a+\frac{1}{2} & b & c+\frac{1}{2} \\ \alpha+\frac{1}{2} & \beta & \gamma-\frac{1}{2} \end{pmatrix}$$
$$= \{\tfrac{1}{2}(a+b+c+1)(-a-b+c) + (a+\tfrac{1}{3})\beta - (b+\tfrac{1}{3})\alpha\} \begin{pmatrix} a & b & c \\ \alpha & \beta & \gamma \end{pmatrix}$$
$$+ \{(a+b+c+1)(a+b-c)(a+\alpha)(b-\beta)\}^{1/2}$$
$$\times \begin{pmatrix} a-\frac{1}{2} & b-\frac{1}{2} & c \\ \alpha-\frac{1}{2} & \beta+\frac{1}{2} & \gamma \end{pmatrix}. \tag{A7}$$

These relations are related to $(ij) = (01), (02), (03), (04), (12)$ and their negative equivalents. The other ones are obtained by circular permutation of (a,b,c) in the coefficients, but restoring the right order in the generalized 3-j symbols; they hold for any values of a,b,c,α,β as long as the square roots can be defined by analytic continuation to real values.

[1]J. Raynal, J. Math. Phys. **19**, 467 (1978).
[2]F.J.W. Whipple, Proc. London Math. Soc. (2) **23**, 104 (1925).
[3]V.A. Knyr, P.P. Pipiraite, and Yu.F. Smirnov, Yad. Fiz. **22**, 1063 (1975).
[4]J. Raynal, Nucl. Phys. A **259**, 272 (1976).
[5]M.S. Kil'dyushov, Yad. Fiz. **15**, 197 (1972).
[6]J. Raynal, Physics Lett. B **45**, 89 (1973).
[7]D.A. Varshalovich, A.N. Moskaliev, and V.K. Khorsonskii, *Quantum Theory of Angular Momenta* (Leningrad, 1975).
[8]F.J.W. Whipple, Proc. London Math. Soc. (2) **40**, 336 (1936).
[9]A. D'Adda, R. D'Auria, and G. Ponzano, Nuovo Cimento, A **23**, 69 (1974).
[10]A.P. Yutsis and A.A. Bandzaitis, *The Theory of Angular Momenta in Quantum Mechanics* (Vilnius, 1965).
[11]T. Regge, Nuovo Cimento, **11**, 116 (1959).
[12]L.A. Shelepin, Sov. Phys. JETP **21**, 228 (1965).
[13]M.E. Rose, *Multipole fields* (Wiley, New York, 1955), p. 93.
[14]B.M. Minton, J. Math. Phys. **11**, 3061 (1970).
[15]K. Srinivasa Rao, T.S. Santhaman, and K. Venkatesh, J. Math. Phys. **16**, 1528 (1975).
[16]L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966).
[17]K. Srinivasa Rao, and K. Venkatesh, *Group Theoretical Methods in Physics*, Proc. of the 5th International Colloquium University of Montreal-July 1976. Edited by R.J. Sharp and B. Kolman (Academic, New York, 1977), p. 649.
[18]G. Racah, Phys. Rev. **62**, 438 (1942).

Meromorphic solutions of nonlinear partial differential equations and many-particle completely integrable systems [a)]

D. V. Chudnovsky

Department of Mathematics, Columbia University, New York, New York 10027
and Centre de Mathematiques, Ecole Polytechnique, Plato de Palaiseau, 91128 Palaiseau Cedex, France
(Received 12 May 1978)

Complete description of meromorphic solutions of several two-dimensional equations with algebraic laws of conservation is obtained. Among them are Zakharov–Shabat systems and, e.g., the Kadomtsev–Petiashvili equation.

INTRODUCTION

We shall consider two-dimensional equations and their meromorphic solutions, especially elliptic solutions (i.e., solutions expressed in terms of elliptic functions). We shall investigate the behavior of these solutions using the picture of poles in the complex plane, so transferring our problem to the many-particles one, in the spirit of Refs. 1 and 2. We will describe completely meromorphic (in particular, rational) solutions to some two-dimensional systems.

Two-dimensional Lax (or Zakharov–Shabat) equations for the vector function $\bar{u}(x,y,t)$ have the form[3]

$$\left[L_1 - \frac{\partial}{\partial t}, L_2 - \frac{\partial}{\partial y}\right] = 0.$$

This is the condition of commutation for two operators acting on functions of x:

$$\frac{\partial L_2}{\partial t} - \frac{\partial L_1}{\partial y} = [L_1, L_2]. \qquad \text{(2 dim)}$$

This equation is the system of equations on coefficients of operators L_1, L_2.[3] For such kind of equations, Zakharov and Shabat have constructed the "algebraic inverse scattering method."

The first nontrivial example of a two-dimensional "inverse scattering integrable" equation is the two-dimensional KdV or Kadomtsev–Petiashvili equation[3,4]:

$$\frac{\partial}{\partial x}(u_t + 6uu_x + u_{xxx}) = -3\alpha\frac{\partial^2}{\partial y^2}u \qquad \text{(2 KdV)}$$

for $u = u(x,y,t)$ and $\alpha = \pm 1$.

What is the method of solving (2 dim) algebraically? We consider some auxiliary stationary problem

$$[L_1, Q] = 0 \qquad \text{(S)}$$

or $(\delta/\delta u)I_p = 0$ for some functional $I_p = \int p\,dx$, such that the stationary $(\partial\bar{u}/\partial t = \partial\bar{u}/\partial y = 0)$ manifold (S) is invariant for the system (2 dim). All the solutions of (S) can be found algebraically, when the orders of L_1 and Q are relatively prime.

Then for any solution $\bar{u}(x) = \bar{u}(x,0,0)$ of (S) we find an evolution in y and t according to (2 dim), because for invariant (S), $u(x,y_0,t_0)$ is a solution of (S) for any y_0, t_0. The formulas of Ref. 5 then allow us to write explicit expressions for $\bar{u}(x,t,y)$ using θ—functions of an algebraic curves.

This method can give only meromorphic solutions. Thus it is much more natural to examine all meromorphic solutions $\bar{u}(x,t,y)$, considering, as in Refs. 1 and 2, the motion of the poles $a_i = a_i(t,y)$ of $\bar{u}(x,t,y)$ in all of the complex x plane.

I. MANY-PARTICLE SYSTEMS CONNECTED WITH MEROMORPHIC SOLUTIONS

Many-particle problems which arise from the evolution of poles of the meromorphic solutions of completely integrable systems are connected with systems of finitely (or infinitely) many particles interacting via the potential $G\mathfrak{P}[x]$ [where $\mathfrak{P}(x)$ is a Weierstrass elliptic function[6]]. In the degenerate case we obtain a system of particles $x_i = x_i(t)$ which interact via the Jacobi potential x^{-2}.[7,8] Thus a Hamiltonian of the form

$$H_{\mathfrak{P}} = \tfrac{1}{2}\sum_{i\in I}\dot{x}_i^2 + G\sum_{i\neq j}\mathfrak{P}(x_i - x_j) \quad \text{occurs.}$$

For a finite I, Moser[7] for the case $\mathfrak{P}(x) = x^{-2}$ and Calogero[8] for an arbitrary $\mathfrak{P}(x)$, have proved that the corresponding Hamiltonian system H possesses Lax representation

$$\frac{dL}{dt} = [A,L].$$

The matrix $L = (L_{ij})$, $i,j\in I$ has the form

$$L_{ij} = (1-\delta_{ij})\sqrt{-G}\,\alpha(x_i - x_j) + \delta_{ij}\dot{x}_i,$$

where $\alpha^2(x) = \mathfrak{P}(x)$.

Then the quantities $J_n = (1/n)\mathrm{tr}(L^n)$, $n\geqslant 1$ are the first integrals of $H_{\mathfrak{P}}$. Moreover it is proved that the J_n are involutive and that they are sums of polynomials in $\dot{x}_i$, $\mathfrak{P}(x_i - x_j)$, G with rational coefficients. The form of the first terms of J_n is the following,

$$J_n = \frac{1}{n}\sum_{i\in I}\dot{x}_i^n + G\sum_{i\neq j}(\dot{x}_i^{n-2} + \dot{x}_i^{n-3}\dot{x}_j + \cdots + \dot{x}_j^{n-2}) \times\mathfrak{P}(x_i - x_j) + \cdots.$$

For the Hamiltonian $H_{\mathfrak{P}}$ with $\mathfrak{P}(x) = x^{-2}$, the exact formulas for solutions can be given for finite $|I| = n$.

For the matrix

$$M(t) = \mathrm{diag}(x_1(t_0),\ldots,x_n(t_0)) + L(t_0)(t - t_0)$$

[a)] Publication of this article was delayed 3 months through no fault of the author.

 0022-2488/79/122416-07$01.00

the eigenvalues $x_i(t)$ are solutions with given initial values $x_i(t_0)$, $\dot{x}_i(t_0)$, corresponding to the Hamiltonian $H = \frac{1}{2}\mathrm{tr}L^2$. For the Hamiltonian $J_n = (1/n)\mathrm{tr}L^n$, L must be replaced by L^{n-1} in $M(t)$.

There is a close relationship between many-particle systems H, J_n, and the solutions of known completely integrable systems.

This connection lies in the method of the so-called pole interpretation first proposed by M. Kruskal.

The ideas of this method are the following: to consider solutions $u(x,t)$ of nonlinear partial differential equations as meromorphic functions in the complex x plane, and to investigate the motion of the poles $x_i = x_i(t)$ as particles with self-consistent potential.

The application of the pole interpretation and the establishment of a connection with the Hamiltonian $H_{\mathfrak{P}}$ can be described along the lines of the following general scheme.[6]

We consider the following special class of meromorphic functions, residues and poles of which are expressed in terms of the variables $(x_i, \dot{x}_i)$ of $H_{\mathfrak{P}}$,

$$\hat{u}_k(z,t) = \sum_{i\in I} \frac{\partial J_{k+1}}{\partial \dot{x}_i} \mathfrak{P}(z - x_i), \quad k = 0,1,2,\cdots. \qquad (O)$$

A special sequence of differential equations connected with (O) exists.

Theorem: Let

$$u_{\underbrace{k,x,\ldots,x}_{(m\text{factors})}}$$

have the weight $k + m + 2$. Then there exist polynomials $\Omega_k(u_0,\ldots,u_{k-1})$ in $u_0, u_{0,x},\ldots,u_1,u_{1x},\ldots,u_{k-1},\ldots,u_{k-1,xx\ldots},\cdots$ of degree two and having all the monomials of weight $k+3$ such that the system of equations

$$u_{k,t} + u_{k+1,x} + \frac{d}{dx}\Omega_k(u_0,\ldots,u_{k-1}) = 0, \quad k = 0,1,\cdots \qquad (C)_k$$

satisfies the following properties:

(1) the functions $\hat{u}_k$ satisfy (C) if and only if $x_i = x_i(t)$ move according to $H_{\mathfrak{P}}$;

(2) if $u_k(x,t)$ satisfy (C) and are meromorphic functions with poles of order 2, then $u_k = \hat{u}_k$.

Here are the first few Ω_k:

$$\Omega_0 = 0, \quad \Omega_1 = -\frac{G}{2}u_0^2 + \frac{G}{12}u_{0xx},$$

$$\Omega_2 = -Gu_0u_1 + \frac{G}{6}u_{1xx},$$

$$\Omega_3 = -\frac{G}{2}u_1^2 - Gu_0u_2 - \frac{G^2}{8}u_{0x}^2 - \frac{G^2}{12}u_0u_{0xx} + \frac{G^2}{120}u_{0xxxx} + \frac{G}{4}u_{2xx}, \quad \cdots.$$

The first such nontrivial system coincides with the Boussinesq equation $u_{tt} + (u^2)_{xx} + u_{xxxx} = 0$. In general, the system (C) describes one of the scheme of approximations of the two-dimensional shallow water equation.

The system (C) also describes many two-dimensional, completely integrable systems, whose poles move according to Hamiltonians J_n. For this we present the system $(C)_0$–$(C)_n$ in the following form:

$$(C)_k : k = 0,\ldots,n-1$$

$$u_{n,t} - u_{0,y} + \frac{d}{dx}\Omega_n(u_0,\ldots,u_{n-1}) = 0. \qquad (\hat{C})$$

Here we replace $u_{n+1,x}$ by $-u_{0,y}$ in (C). Then it follows from the theorem that the functions $\hat{u}_k : k = 0,\ldots,n$ from (O) satisfy ($\hat{C}$) if and only if x_i move in the t direction according to $H_{\mathfrak{P}} = J_2$ and in the y direction according to J_n.

II. MEROMORPHIC SOLUTIONS OF TWO-DIMENSIONAL EQUATIONS AND THEIR POLES

We already know by Refs. 1, 2, and 9 that the evolution of poles of several one-dimensional equations (e.g., KdV, Boussinesq,···) is connected with the Hamiltonians $H_{\mathfrak{P}}$ and $J_n = (1/n)\mathrm{tr}(L^n)$, $n = 2,3,\cdots$ of Refs. 1, 2, and Sec. I.

These one-dimensional systems have Lax' form

$$\frac{dL_1}{dt} = [L_1, L_2],$$

and so are included in the more general two-dimensional Eq. (2 dim). In this paper we shall consider some natural conjectures about meromorphic solutions of (2 dim) and obtain new solutions of (2 KdV) and similar equations in terms of elliptic functions.

The chain (C) of nonlinear equations with constraints $u_n \equiv 0$ gives some particular system of nonlinear equations of evolution possessing infinitely many algebraic laws of conservation. As we had already mentioned in Sec. I the first system in this chain is the Boussinesq equation, i.e., corresponding to the Lax pair $dA/dt = [L,A]$, where L is Schrodinger and A is of degree 3. For this equation, the motion of the poles $x_i = x_i(t)$ corresponds to the motion according to the Hamiltonian $H = J_2$ with the restrictions $\mathrm{grad}J_3 \equiv 0$.[3] On the other hand, the KdV equation has the Lax representation $dL/dt = [A,L]$ and the poles $x_i = x_i(t)$ move according to the Hamiltonian J_3 with the restrictions $\mathrm{grad}H = \mathrm{grad}J_2 = 0$ (see Refs. 1 and 2).

So it is natural to put out the following conjecture.

Conjecture: The pole evolution of the system having the Lax form $dL/dt = [A,L]$ for L of order n and A of order m, $n > 1$, $m > 1$, is connected with the system with the Hamiltonian J_m with the restriction $\mathrm{grad}J_n = 0$. By analogy, if we have the system $dL/dt - dA/dy = [A,L]$, then the poles $x_i = x_i(y,t)$ move in y according to J_n and in t according to J_m.

This conjecture was proved in Ref. 1 for the case $m = 2$, $n = 3$ of $m = 3$, $n = 2$ (see Ref. 1, §10, p. 350). In Ref. 10 it is mentioned that these results were obtained "very recently" (after Ref. 1). We also prove here the conjecture for $n = 2$, $m = 4$ and show its close relation with system (C).

A. Meromorphic, rational, and elliptic solutions of the two-dimensional Korteweg–de Vries equation

In this part, general meromorphic solutions of (2 KdV) are considered and new elliptic solutions are constructed.

1. Meromorphic solutions

For the two-dimensional Korteweg–de Vries equation

$$\frac{\partial}{\partial x}(u_t + 6uu_x + u_{xxx}) = -3\alpha \frac{\partial^2}{\partial y^2} u, \qquad (2\text{ KdV})$$

we shall consider the meromorphic solutions, written in the general form

$$u(x,y,t) = \sum_{i\in I}(-2)(x-a_i)^{-2}, \quad a_i = a_i(y,t) \qquad (1)$$

or

$$u(x,y,t) = \sum_{i\in I}(-2)\mathfrak{P}(x-a_i), \quad a_i = a_i(y,t). \qquad (2)$$

It is easy to show that, if $u(x,y,t)$ is a meromorphic solution in x of (2 KdV) for $(y,t)\in[0,y_0]\times[0,t_0]$, then the poles in $a_i(y,t)$ are of second order with residues -2. Thus the form (1) or (2) is the general form of the meromorphic solutions of (2 KdV).

In Ref. 1 (§10, p. 350) it was shown that the motion of the poles $a_i(y,t)$ is in y according to $J_2 \equiv H$ and in t according to J_3. This is possible because J_2 and J_3 commutes. More precisely we have

Proposition 1: The function

$$u(x,y,t) = -2\sum_{i\in I}\mathfrak{P}(x-a_i), \quad a_i = a_i(y,t)$$

is a meromorphic solution of (2 KdV) if and only if

$$\begin{aligned} &\alpha a_{iyy} = 4\sum_{j\neq i}\mathfrak{P}'(a_i - a_j), \\ &a_{it} = 3\alpha a_{iy}^2 - 12\sum_{j\neq i}\mathfrak{P}(a_i - a_j): i\in I. \end{aligned} \qquad (3)$$

Proof: By the law of addition for $\mathfrak{P}(x)$ in the form 13.13(10), Ref. 11, we have

$$\begin{aligned} 6uu_x = {} & 12\sum_{i\in I}(\mathfrak{P}^2)'(x-\alpha_i) + 24\sum_{i\in I}\mathfrak{P}(x-\alpha_i) \\ & \times\sum_{j\neq i}\mathfrak{P}(\alpha_i-\alpha_j) + 24\sum_{i\in I}\mathfrak{P}(x-\alpha_i)\sum_{j\neq i}\mathfrak{P}(\alpha_i-\alpha_j). \end{aligned}$$

Because $\mathfrak{P}'' = 6\mathfrak{P}^2 - g_2/2$ we obtain from (2KdV) the system (3) looking at coefficients at $\mathfrak{P}(x-\alpha_i)$, $\mathfrak{P}'(x-\alpha_i)$, where $\mathfrak{P}(x)\sim x^{-2}$, $\mathfrak{P}'(x)\sim -2x^{-3}$.

If we consider the commuting flows

$$\begin{aligned} &\bar{J}_2 = \sum_{i\in I}\frac{b_i^2}{2} - \frac{4}{\alpha}\sum_{i\neq j}\mathfrak{P}(a_i-a_j), \\ &\bar{J}_3 = \alpha\sum_{i\in I}b_i^3 - \sum_{i\neq j}(b_i+b_j)\mathfrak{P}(a_i-a_j), \end{aligned}$$

i.e., $G = 4/\alpha, \bar{J}_2 = H = J_2$, and $\bar{J}_3 = 3\alpha J_3$, then the system (3) is obviously equivalent to

$$\begin{aligned} &a_{iy} = \frac{\partial \bar{J}_2}{\partial b_i}, \quad b_{iy} = -\frac{\partial \bar{J}_2}{\partial a_i}, \\ &a_{it} = \frac{\partial \bar{J}_3}{\partial b_i}, \quad b_{it} = -\frac{\partial \bar{J}_2}{\partial a_i}, \quad i\in I. \end{aligned} \qquad (4)$$

Because $\bar{J}_2$ and $\bar{J}_3$ are commuting we obtain solutions $a_i(y,t)$, $b_i(y,t)$ of system (4), or of system (3).

Thus starting from any initial data at $y = y_0$, $t = t_0$,

$$a_i^0 = a_i(y_0,t_0), \quad b_0^i = a_{iy} = b_i(y_0,t_0), \quad i\in I,$$

we can explicitly find (at least for finite I) a solution of (3), $a_i(y,t)$ such that

$$a_i(y_0,t_0) = a_i^0, \quad a_{iy}(y_0,t_0) = b_i^0.$$

For example, for rational solutions $u(x,y,t)$ of (2 KdV) of degree $2n$, this gives solutions depending on $2n$ arbitrary parameters.

2. Rational solutions

How do we obtain, e.g., rational solutions of (2 KdV), e.g., those when $I = n$, $\mathfrak{P}(x) = x^{-2}$? According to the theory of Hamiltonians commuting with $H_{x^{-2}}$ described in Refs. 1 and 2, we have the following.

Rule 2: If we have initial conditions at $y = y_0$, $t = t_0$,

$$a_i^0(y_0,t_0) = a_i^0, \quad b_i^0(y_0,t_0) = b_i^0,$$

then for two flows having Hamiltonians

$$H_1 = \operatorname{tr} f(L), \quad H_2 = \operatorname{tr} g(L),$$

for which

$$\begin{aligned} &a_{iy} = \frac{\partial H_1}{\partial b_i}, \quad b_{iy} = -\frac{\partial H_1}{\partial a_i}, \\ &a_{it} = \frac{\partial H_2}{\partial b_i}, \quad b_{it} = -\frac{\partial H_2}{\partial a_i}, \\ &a_i(y_0,t_0) = a_i^0, \quad b_i(y_0,t_0) = b_i^0, \quad i = 1,\ldots,n. \end{aligned} \qquad (5)$$

The solutions are defined as eigenvalues of the following matrix,

$$\begin{aligned} \mathfrak{U}_{H_1,H_2} = {} & \operatorname{diag}(a_1^0,\ldots,a_n^0) + (y-y_0)f'(L)(y_0,t_0) \\ & + (t-t_0)g'(L)(y_0,t_0). \end{aligned}$$

Now it is clear how to obtain $u(x,y,t)$.

Rule 3: If

$$u(x,y,t) = -2\sum_{i=1}^{n}(x-a_i)^{-2}, \quad a_i = a_i(y,t), \quad i = 1,\ldots,n$$

is such that a_i moves in y according to $H_1 = \operatorname{tr} f(L)$ and in t according to $H_2 = \operatorname{tr} g(L)$, then

$$u(x,y,t) = 2\frac{d^2}{dx^2}\ln\chi(x,\mathfrak{U}_{H_1,H_2}),$$

where $\chi(x,\mathfrak{U}_{H_1,H_2})$ is a characteristic polynomial of the matrix $\mathfrak{U}_{H_1,H_2}$ defined before.

Thus for two-dimensional KdV (2 KdV),

$$u(x,y,t) = 2\frac{d^2}{dx^2}\ln\chi(x,\mathfrak{U}) \qquad (6)$$

for $\mathfrak{U} = \mathfrak{U}_{J_2,3\alpha J_3}$,

$$\mathfrak{U} = \operatorname{diag}[a_i^0] + (y-y_0)L(t_0,y_0) + (t-t_0)3\alpha L^2(t_0,y_0). \qquad (7)$$

So all rational solutions of (2 KdV) are easily described.

There exists a paper by Manakov, Zakharov, Bordag, Its, and Matveev[12] where they have written similar formulas for rational solutions with $\alpha = -1$. They deduced these solutions from solutions of the Zakharov–Shabat type with a degenerate kernel taking in $\exp(x_i y + \eta_i t)$ the limit $x_i \to 0$,

$\eta_i \to 0$, i.e., considering the degenerate case of exponential functions.

For example, let us write down the simplest nonsingular rational solution. We also must mention that in Ref. 10 this solution was written incorrectly. It is

$$u(x,y,t) = 2\frac{4/\nu + 2\nu y^2 - 2(x - 3\nu^2 t - x_0)}{[(x - 3\nu^2 t - x_0)^2 + \nu^2 y^2 + 2/\nu]^2}. \quad (8)$$

Here $\alpha = -1$ and $v_x = 3\nu^2$, velocity in the x direction, $v_y = -6\nu$ in the y direction of "soliton."

Now we shall proceed to exhibit very interesting elliptic solutions of (2 KdV). The simplest elliptic solution is of the following form,

$$u(x,y,t) = -2\mathfrak{P}(x + By + (3V - 3\alpha B)t + x_0) + V. \quad (9)$$

We shall now consider the elliptic solution of (2 KdV) with two poles in each part of the lattice. So we take

$$u_2(x,y,t) = -2\mathfrak{P}(x - a_1) - 2\mathfrak{P}(x - a_2), \quad (10)$$

with $a_i = a_i(x,t)$, $i = 1,2$. Then according to the equation of evolution, after some changes, we obtain the following formulas:

$$a_1 = R_0 y + C_1 t + C_2 + \eta(y + 6\alpha R_0 t)/2,$$
$$a_2 = R_0 y + C_1 t + C_2 - \eta(y + 6\alpha R_0 t)/2,$$

where the function $\eta(y + G\alpha R_0 t)$ satisfies

$$3\alpha\eta_y^2/4 = C_1 - 3\alpha R_0^2 + 12\mathfrak{P}(\eta).$$

3. Elliptic solutions

We have already reduced the problem of finding elliptic solutions of two-dimensional equations, especially of (2 KdV) to the solution of ordinary differential equations involving $\mathfrak{P}(z)$, such that

$$\frac{\partial a_i}{\partial y} = \frac{\partial H}{\partial b_i}, \quad \frac{\partial b_i}{\partial y} = -\frac{\partial H}{\partial a_i}$$

for $H = H_{\mathfrak{P}}$.[1,2]

We know that this system for finite $|I|$ is completely integrable. But we do not know what is the exact form of these solutions when $|I| > 2$. Here we have the following questions:

(1) Whether the general trajectory $a_i(y)$, $b_i(y)$ can be expressed using only elliptic functions, namely using $|I| - 1$ additional ones; what are the relations between invariants of these functions and $\mathfrak{P}(x)$, if elliptic representation is possible?

(2) What is topologically the variety of solutions of $H_{\mathfrak{P}}$?

In the case of degenerate $\mathfrak{P}(x)$, e.g., $\mathfrak{P}(x) = x^{-2}$, $\sin^{-2}x$, $\sinh^{-2}x$, we have simple formulas for exact solutions (cf. Ref. 2). We also have the result for $|I| = 2$.

Let

$$a_{1yy} = -G\mathfrak{P}'(a_1 - a_2), \quad a_{2yy} = -G\mathfrak{P}'(a_2 - a_1).$$

Then $a_1 + a_2 = A_0 y + A_1$ and for $a = a_1 - a_2$,

$$a_y^2 = (-4G)\mathfrak{P}(a) - 4GC,$$

C is a constant. If we put

$$a(y) = a(y_1/2\sqrt{-G})$$

and set

$$v = \mathfrak{P}(a),$$

then for

$$\xi = -(7C^3 - g_2C/4 - g_3/4)[\mathfrak{P}(a) + C]^{-1} + 3C^2 - g_2/12,$$

we have the elliptic representation for ξ, ξ satisfying

$$\xi_{y_1}^2 = 4\xi^2 - G_2\xi - G_3,$$

where

$$G_2 = -3g_3C + g_2^2/12 + g_2C^2,$$
$$G_3 = g_3C^3 - g_2^2C^2/6 + g_2g_3C/4 - g_3^2/4 + g_2^3/6^3.$$

If $\mathfrak{P}_1(z)$ is an elliptic function, satisfying

$$\mathfrak{P}_{1z}^2 = 4\mathfrak{P}_1^3 - G_2\mathfrak{P}_1 - G_3,$$

then

$$\xi = \mathfrak{P}_1(2\sqrt{-G}\,y + y_0).$$

It is very important to mention that for a modular invariant $j = j(\mathfrak{P})$,

$$j = g_2^3/(g_2^3 - 27g_3^2), \quad \text{if } \Delta \neq 0,$$

the modular invariant $J = J(\mathfrak{P}_1)$ can be chosen (with variation of C) *arbitrarily*. Thus the complete solution of $H_{\mathfrak{P}}$ involves two functions, the $\mathfrak{P}(z)$ and a new $\mathfrak{P}_1(z)$ with arbitrary invariant.

4. An elliptic 2-soliton

We turn back to the solution (10) of (2 KdV). Now we use the solution of $H_{\mathfrak{P}}$ for $|I| = 2$. We put

$$\xi(y + 6\alpha R_0 t) = -\left(7C^3 - \frac{g_2}{4}C - \frac{g_3}{4}\right) \times 1/[\mathfrak{P}(\eta(y + 6\alpha R_0 t)) + C] + 3C^2 - \frac{g_2}{12},$$

where

$$C = \frac{C_1}{12} - \frac{R_0^2\alpha}{4}.$$

Then we have

$$\xi(y + 6\alpha R_0 t) = \mathfrak{P}_1(4(y + 6\alpha R_0 t)/\sqrt{\alpha} + C_3). \quad (11)$$

Here the function $\mathfrak{P}_1$ satisfy $\mathfrak{P}_1'^2 = 4\mathfrak{P}_1^3 - G_2\mathfrak{P}_1 - G_3$, and G_2, G_3 were defined before. This gives exact formulas for $u_2(x,y,t)$ in (10), depending on $g_2, g_3, C_1, C_2, R_0, C_3$. As we already mentioned before this leads to *nonsimple* Abelian variety

$$(\mathfrak{P}(x), \mathfrak{P}'(x), \mathfrak{P}_1(x), \mathfrak{P}_1'(x))$$

of dimension two and *any* nonsimple Abelian variety of dimension two can appear as a manifold for solutions of (2 KdV).

The solutions $u_2(x,y,t)$ are nonsingular and especially interesting if $\mathfrak{P}(z)$ is with complex multiplication on $\sqrt{-1}$, i.e., Ref. 11, $g_3 = 0$. In this case the a_i can be chosen such that $a_1 = \sqrt{-1}\,a$, $a_2 = -\sqrt{-1}\,a$, and $u_2(x,y,t)$ is bounded in the (x,y) plane.

B. Meromorphic solutions of other two-dimensional equations

The very general idea of pole expansion can be used in order to obtain information even about algebraic properties of equation. We shall use the pole expansion in this part for different two-dimensional systems of the form (2 dim) with L_1, L_2 of order two and four.

1. Elliptic solutions for (2 dim) in the case $n = 2$, $m = 4$

The (2 KdV) corresponds to the case $dL_2/dt - dL_3/dy = [L_2, L_3]$ and we get a good description in terms of Hamiltonians J_2, J_3. The next case (a very nontrivial one) is the system

$$\frac{dL_2}{dy} - \frac{dL_4}{dt} = [L_2, L_4],$$

giving equations for the coefficients of the operators of degree two, L_2 and four, L_4. If we consider them as usual,

$$L_2 = \frac{d^2}{dx_2} + u, \quad L_4 = \frac{d^4}{dx^4} + u_2\frac{d^2}{dx^2} + u_1\frac{d}{dx} + u_0,$$

then we have the following system for u_0, u_1, u_2:

$$\begin{aligned} &u = u_2/2 \\ &u_{2t} = 2u_{2xx} - 2u_{1x}, \\ &u_{1t} = -u_{1xx} + 2u_{2xxx} + u_2u_{2x} - 2u_{0x}, \\ &u_{0t} - \frac{u_{2y}}{4} u_{2xxxx} + \frac{1}{2}u_2u_{2xx} - u_{0xx} + \frac{1}{2}u_{2x}u_1. \end{aligned} \tag{L}$$

We can transform this system into a more convenient form.

If we put

$$u_1 = u_{2x} + \hat{u}_1/2, \quad u_0 = \tfrac{1}{3}u_{2xx} + \hat{u}_{1x}/4 + \hat{u}_0/4,$$

then the system (L) takes the form

$$\begin{aligned} &u_{2t} = -\hat{u}_{1x}, \\ &\hat{u}_{1t} = \tfrac{2}{3}u_{2xxx} + 2u_2u_{2x} - \hat{u}_{0x}, \\ &\hat{u}_{0t} - 2u_{2y} = \tfrac{1}{3}\hat{u}_{1xxx} + \hat{u}_1u_{2x}. \end{aligned} \tag{$\hat{\text{L}}$}$$

Now we consider the general *purely Weierstrass elliptic* solution of ($\hat{\text{L}}$). If the functions $\hat{u}_0, \hat{u}_1, u_2$ are purely Weierstrass elliptic solutions of (L), they can have the form

$$u_2 = -4\sum_{i\in I}\mathfrak{P}(x - a_i); \tag{12}$$

$$\hat{u}_1 = -4\sum_{i\in I}a_{it}\mathfrak{P}(x - a_i); \tag{13}$$

$$\hat{u}_0 = \sum_{i\in I}\left[-4a_{it}^2 + 32\sum_{j\neq i}\mathfrak{P}(a_i - a_j)\right]\mathfrak{P}(x - a_i). \tag{14}$$

We must mention that (12)–(14) is not the most general form of meromorphic and even elliptic solutions of (L), but it is the general form for purely Weierstrass elliptic solutions.

In order for (12)–(14) to satisfy ($\hat{\text{L}}$), the following condition must be satisfied,

$$a_{itt} = 8\sum_{j\neq i}\mathfrak{P}'(a_i - a_j), \quad i\in I. \tag{15}$$

However, it can easily be shown from ($\hat{\text{L}}$) that (12)–(14) satisfy ($\hat{\text{L}}$),

if $a_{it} = a_{jt}$ for all $i, j\in I$.

Thus, for all the elliptic solutions of ($\hat{\text{L}}$) the velocity of poles in the t direction is the same for all the poles.

The conditions for satisfaction of ($\hat{\text{L}}$) for (12)–(14) are the following:

$$\sum_{j\neq i}\mathfrak{P}'(a_i - a_j) = 0, \quad i\in I, \quad a_{it} = V, i\in I.$$

This is the condition $\operatorname{grad}(J_2 - VJ_1) = 0$. In the y direction, a_i moves according to J_4 but linearly in t.

Even for y independent (L) this gives N-soliton solutions which have the same velocity not connected with initial conditions. The solution thus remains always connected solitons with a strange interaction of the 1-solitons inside the N-soliton. The simplest 1-soliton of (L), y independent, is the following, obtained first by Manin[13]:

$$\begin{aligned} \mu_1 &= 2a^2\frac{\cosh 2ax - \cos 2ax - 2\sin 4a^2t}{\cosh 2ax + \cos 2ax + 2\cos 4a^2t}, \\ \mu_2 &= -2a\frac{\sinh 2ax - \sin 2ax}{\cosh 2ax + \cos 2ax + 2\cos 4a^2t}, \\ u_2 &= -4\mu_{2x}, \quad u_1 = -6\mu_{2xx} - 4\mu_{1x} + 4\mu_2\mu_{2x}, \\ \mu_0 &= -4\mu_{2xxx} - 6\mu_{1xx} + 8(\mu_{2x})^2 + 4(\mu_1\mu_2)_x \\ &\quad + 6\mu_2\mu_{2xx} - 4\mu_{2x}\mu_2^2, \end{aligned}$$

or we change x by $x + c$. These solutions have the period $\pi/2a^2$, but when $\cos 4a^2t = -1$, we have "explosion" at $x = 0$.

2. New systems of equations connected with integrals J_2 and J_4

It is clear that because of the strange behavior of the elliptic Weierstrass solutions of (L), even for the y-independent case, it is necessary to correct (L).

We correct (L) using the system of equations $(\hat{\text{C}})_0$–$(\hat{\text{C}})_2$ from Sec. I in such a way that it has good elliptic (as well as meromorphic) solutions with poles moving according to J_4 in the y direction and according to J_2 in t.

Here is the system ($\hat{\text{C}}$):

$$\begin{aligned} &u_{2t} = -\hat{u}_{1x}, \\ &\hat{u}_{1t} = -\hat{u}_{0x} - (G/12)u_{2xxx} - (G/4)u_2u_{2x}, \\ &\hat{u}_{0t} - u_{2y} = -(G/6)\hat{u}_{1xxx} - (G/4)u_{2x}\hat{u}_1 - (G/4)u_2\hat{u}_{1x}. \end{aligned} \tag{16}$$

The meromorphic solutions of this system are of the same type:

$$\hat{u}_2 = -4\sum_{i\in I}\mathfrak{P}(x - a_i);$$

$$\hat{u}_1 = -4\sum_{i\in I}a_{it}\mathfrak{P}(x - a_i);$$

$$\hat{u}_0 = -4\sum_{i\in I}\left[a_{it}^2 + G\sum_{j\neq i}\mathfrak{P}(a_i - a_j)\right]\mathfrak{P}(x - a_i),$$

but the a_i moves in y according to J_4 and in t according to J_2, for any G.

Now we will show that the systems ($\hat{\text{L}}$) and its correct form (16) are in fact equivalent. So it is possible to obtain an

analog of the elliptic solutions even for $(\hat{L})$.

To do this, we will first investigate the general form of meromorphic solutions for $(\hat{L})$, not necessarily pure Weierstrass elliptic.

3. General meromorphic solutions of (L) are connected with J_2 and J_4

The general meromorphic solutions of $(\hat{L})$ are really connected with the Hamiltonians J_2 and J_4, but not with $G=-8$ (as for pure Weierstrass elliptic solutions) and with $G=-4$. However, in this case, u_0 contains the poles of first order.

Now let us find *all* meromorphic solutions of $(\hat{L})$; they have rather a simple nature, connected with J_2 and J_4. We must have, for an arbitrary meromorphic solution $\hat{u}_0, \hat{u}_1, u_2$ of $(\hat{L})$, the following representation:

$$u_2 = -4\sum_{i\in I}(x-a_i)^{-2}, \tag{17}$$

$$u_1 = -4\sum_{i\in I}a_{it}(x-a_i)^{-2}, \tag{18}$$

$$\hat{u}_0 = -4\sum_{i\in I}\left[a_{it}^2 - 8\sum_{j\neq i}(a_i-a_j)^{-2}\right](x-a_i)^{-2} - 4\sum_{i\in I}\left[a_{itt} + 16\sum_{j\neq i}(a_i-a_j)^{-3}\right](x-a_i)^{-1}. \tag{19}$$

The general functions $u_2, \hat{u}_1, \hat{u}_0$ (17)–(19), as functions from x, t satisfy $(\hat{L})$ if

$$a_{itt} = -8\sum_{j\neq 1}(a_i-a_j)^{-3}, \quad i\in I, \tag{20}$$

i.e., a_i as functions of t satisfy H_{x-2} for $G=-4$. Similarly $a_i(y,t)$ as a function of y is the solution corresponding to J_4, also for $\mathfrak{P}(x)=x^{-2}$ and $G=-4$.

Thus, rational solutions of (L) really satisfy our conjecture.

Now we will write $(\hat{L})$ in a more traditional form [cf. (C) in Sec. I and the expressions for $\hat{u}_k(x;t)$ in Sec. I]. To do this, we remember that (17)–(19) is the solution of $(\hat{L})$ if (20) is satisfied,

$$a_{itt} = -8\sum_{j\neq i}(a_i-a_j)^{-3}.$$

Thus

$$\hat{u}_0 = -4\sum_{i\in I}\left[a_{it}^2 - 8\sum_{j\neq i}(a_i-a_j)^{-2}\right](x-a_i)^{-2} - 4\sum_{i\in I}\left[8\sum_{j\neq i}(a_i-a_j)^{-3}\right](x-a_i)^{-1}.$$

We put

$$\tilde{\hat{u}}_0 = -4\sum_{i\in I}\left[a_{it}^2 - 4\sum_{j\neq i}(a_i-a_j)^{-2}\right](x-a_i)^{-2}. \tag{21}$$

From (21) and (17)–(19) it follows that

$$\hat{u}_0 = \tilde{\hat{u}} - 4\sum_{i\in I}\left[-4\sum_{j\neq i}(a_i-a_j)^{-2}\right](x-a_i)^{-2} - 4\sum_{i\in I}\left[8\sum_{j\neq i}(a_i-a_j)^{-3}\right](x-a_i)^{-1}.$$

Now for $\mathfrak{P}(x)=x^{-2}$ we have

$$(\hat{u}_0-\tilde{\hat{u}}_0)_x = -4\sum_{i\in I}\left[-4\sum_{j\neq i}\mathfrak{P}(a_i-a_j)\right]\mathfrak{P}'(x-a_i) + 4\sum_{i\in I}\left[-4\sum_{j\neq i}\mathfrak{P}'(a_i-a_j)\right]\mathfrak{P}(x-a_i).$$

By the law of addition for $\mathfrak{P}(x)$ we have

$$(\hat{u}_0-\tilde{\hat{u}}_0)_x = u_2u_{2x} - 16\sum_{i\in I}\mathfrak{P}(x-a_i)\mathfrak{P}'(x-a_i) = u_2u_{2x} + \tfrac{1}{3}u_{2xxx}.$$

Thus,

$$\hat{u}_{0x} = \tilde{\hat{u}}_{0x} + u_2u_{2x} + \tfrac{1}{3}u_{2xxx}.$$

Now we can rewrite $(\hat{L})$ in the following form:

$$u_{2t} = -\hat{u}_{1x},$$

$$\hat{u}_{1t} = \tfrac{1}{3}u_{2xxx} + u_2u_{2x} - \tilde{\hat{u}}_{0x},$$

$$\tilde{\hat{u}}_{0xt} - 2u_{2xy} = \tfrac{1}{3}\hat{u}_{1xxx} + \hat{u}_{1x}u_{2x} + \hat{u}_1u_{2xx} - u_{2t}u_{2x} - u_2u_{2xt} - \tfrac{1}{3}u_{2txx} = \tfrac{2}{3}\hat{u}_{1xxx} + 2\hat{u}_{1x}u_{2x} + \hat{u}_1u_{2xx} + \hat{u}_{1xx}u_2 = \left(\tfrac{2}{3}\hat{u}_{1xx} + (\hat{u}_1u_2)_x\right)_x.$$

Now we define $\bar{u}_0(x,t)$ by

$$\bar{u}_0 = \hat{u}_0 - \tfrac{1}{2}u_2^2 - \tfrac{1}{3}u_{2xx}.$$

Then from $(\hat{L})$ we obtain

$$u_{2t} = -\hat{u}_{1x},$$

$$\hat{u}_{1t} = -\bar{u}_{0x} + \tfrac{1}{3}u_{2xxx} + u_2u_{2x}, \tag{$\bar{L}$}$$

$$\bar{u}_{0t} - 2u_{2y} = \tfrac{2}{3}\hat{u}_{1xxx} + \hat{u}_{1x}u_2 + \hat{u}_1u_{2x}.$$

Then, in view of the given supra, this system is equivalent to $(\hat{L})$.

General meromorphic solutions of $(\bar{L})$ have the form (17), (18), (21), with $\tilde{\hat{u}}_0 = \bar{u}_0$, where (20) is satisfied and $a_i(y,t)$ moves according to J_2 on t and according to J_4 on y.

But the system $(\bar{L})$ is equivalent also to (16) by B2, if we change y by $2y$.

Thus, after some transformations, we find that (L) has elliptic solutions and in this case also, the conjecture is true.

CONCLUSION

System (C) gives us the possibility to verify the conjecture for $\min\{n,m\}=2$. For $\min\{n,m\}>2$ besides motion corresponding to J_n and J_m there can appear new many-particle systems.

[1] D.V. Chudnovsky and G.V. Chudnovsky, Nuovo Cimento B **40**, 339–53 (1977).
[2] D.V. Chudnovsky, Seminaire sur les equations nonlineares, I, 1977–1079. Ecole Polytechnique (Expose 1,2,3).
[3] V.E. Zakharov and A.B. Shabat, Funct. Anal. Appl. **8**, 43 (1974).
[4] B.B. Kadomtsev and V.I. Petiashvili, Sov. Phys. Doklady 15, 539 (1970).
[5] J.L. Burchnall and T.W. Chaundy, Proc. London Math. Soc. **21**, 420–40 (1922); Proc. Royal Soc. London, Ser. A **118**, 557–73 (1928); **134**, 471 (1931).
[6] D.V. Chudnovsky, Not. AMS **24**, A-387 (1977).

[7]J. Moser, Adv. Math. **16**, 197 (1975).
[8]F. Calogero, Lett. Nuovo Cimento **13**, 411 (1975).
[9]H. Airault, H.P. McKean, and J. Moser, Commun. Pure Appl. Math. **55**, 1–75 (1977).
[10]I. Krishiver, Usp. Math. Nauk **32**, No. 6 (1977), in Russian.
[11]A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. 2.
[12]S.V. Manakov, V.E. Zakharov, L.A. Bordag, A.R. Its, and V.B. Matveev, Phys. Lett. A **63**, 205 (1977).
[13]Y. Manin, Funct. Anal. Appl. (to appear).

A particular N-soliton solution and scalar wave equations[a)]

W. E. Couch and R. J. Torrence
Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta

(Received 5 June 1979; accepted for publication 21 August 1979)

A special class of N-soliton solutions of the Korteweg–deVries equation is examined. It is shown that the nonscattering wave equation based on the corresponding special reflectionless potentials can be obtained by a coordinate transformation of the ordinary wave equation and that these reflectionless potentials and others occur in the scalar wave equation on black hole geometries of general relativity.

1. INTRODUCTION

It is well known that the N-soliton solutions[1,2] of the Korteweg–deVries (KdV) equation may be obtained from reflectionless potentials of the Schrödinger equation.[3] Among these potentials there is one class which is simpler than the rest. In this paper, we show that this class of potentials is connected in an unexpected manner to the wave equation in Minkowski space and that, in addition, it is related to the wave equation in important geometries of general relativity.

2. A SPECIAL N-SOLITON SOLUTION

A remarkable technique for finding solutions of the KdV equation is the inverse scattering method,[3] in which a solution $u(x,t)$ is generated from its initial data $u(x,0)$ in a series of steps, the first of which is the solving of the time-independent Schrödinger equation with potential $u(x,0)$, and energy λ

$$\psi'' + [\lambda - u(x,0)]\,\psi = 0. \tag{1}$$

The N-soliton solutions are precisely those $u(x,t)$ generated by those $u(x,0)$ which are nonsingular *reflectionless* potentials. It was shown by Kay and Moses[4] that the set of all nonsingular reflectionless potentials for (1) is given by

$$u(x,0) = -2\frac{d^2}{dx^2}\ln\det[\mathscr{I} + C(x,0)], \tag{2}$$

where $\mathscr{I}$ is the $N\times N$ identity matrix and

$$C(x,0) = \left[c_n c_m\, e^{-[(\kappa_n+\kappa_m)x]}/(\kappa_n+\kappa_m)\right], \tag{3}$$

with the c_n positive constants and the κ_n distinct positive constants. The resulting (N-soliton) solutions of the KdV equation are given by

$$u(x,t) = -2\frac{\partial^2}{\partial x^2}\ln\det[\mathscr{I} + C(x,t)], \tag{4}$$

where

$$C(x,t) = \left[c_n c_m \frac{\exp[-(\kappa_n x - \kappa_n^3 t) - (\kappa_m x - \kappa_m^3 t)]}{\kappa_n + \kappa_m}\right].$$

In a study of linear hyperbolic partial differential equations whose solutions have "the characteristic propagation property," i.e., are nonscattering, Kundt and Newman[5] achieved both less and more than did Kay and Moses in presenting reflectionless potentials for (1). On the one hand, they did not obtain the full set of $u(x,0)$ associated with (2) and (3); on the other hand, they did obtain a discrete infinity of such potentials of the simple form

$$u(x,0) = -l(l+1)/\cosh^2 x, \quad l \text{ an integer}. \tag{5}$$

Since these are nonsingular functions, they must be among the potentials of Kay and Moses, i.e., there must be a choice of N and of c_n, κ_n, $n = 1,\cdots, N$, such that

$$-2\frac{d^2}{dx^2}\ln\det[\mathscr{I} + C(x,0)] = \frac{-l(l+1)}{\cosh^2 x}. \tag{6}$$

We shall find that choice.

It is known[6] that the κ_n, $n = 1,\cdots,N$, are related to the eigenvalues of (1) by $\lambda = -\kappa_n^2$, and that[7] the eigenvalues of (1) for the potential given by (5) are $\lambda = -n^2$, $n = 1,\cdots,l$. It follows that $N = l$ and $\kappa_n = n$, $n = 1,2,\cdots,l$. We shall now show that the c_n are given by

$$c_n^2 = \frac{(l+n)!}{(l-n)!n!(n-1)!}. \tag{7}$$

Eigenfunctions of (1), with (5) as potential, are $\phi_n = e^{-nx}F_n$, where F_n is the hypergeometric function

$$F_n = n!\sum_{m=0}^{l}\frac{(-1)^m(l+m)!r^m}{m!(l-m)!(l+n)!}, \tag{8}$$

with r defined by

$$r = (1+e^{2x})^{-1}.$$

The c_n are normalization constants[6] and are determined by

$$c_n^2\int_{-\infty}^{\infty}\phi_n^2\,dx = c_n^2\int_{-\infty}^{\infty}e^{-2nx}F_n^2\,dx = 1. \tag{9}$$

This integral is most easily evaluated in terms of the variable r, for which we have

$$c_n^{-2} = \tfrac{1}{2}\int_0^1 r^{n-1}(1-r)^{-n-1}F_n^2\,dr. \tag{10}$$

By using the product rule of differentiation,

$$D^{l-n}[r^l(1-r)^l] = l!l!(l-n)!(-1)^{l-n}\times\sum_{i=0}^{l-n}\frac{(-1)^i r^{l-i}(1-r)^{i+n}}{(l-n-i)!i!(l-i)!(n+i)!}, \tag{11}$$

where $D \equiv d/dr$, it is easily seen that the following identity holds:

a)Work supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/79/122423-04$01.00

$$F_n = (n!/l!)\, r^{-n}\, D^{l-n}\, G, \tag{12}$$

where $G \equiv r^l(1-r)^l$. It follows from (12) that the integrand in (10) is a polynomial in r and hence the integral is explicitly seen to be well defined.

Upon substitution of (12) into (10) we obtain

$$c_n^{-2} = \frac{1}{2}\left(\frac{n!}{l!}\right)^2 \int_0^1 r^{-n-1}(1-r)^{-n-1}[D^{l-n}\, G\,]^2\, dr. \tag{13}$$

The cases $n = l$ and $n = l-1$ may be obtained directly from (13) and the formula

$$\int_0^1 r^\alpha\,(1-r)^\beta\, dr = \frac{\alpha!\beta!}{(\alpha+\beta+1)!}. \tag{14}$$

Consider the remaining cases $1 \leqslant n \leqslant l-2$ and expand *one* of the factors of $D^{l-n}G$ in (13) in the form given by (11). This gives

$$\begin{aligned} r^{-n-1}(1-r)^{-n-1}\, D^{l-n}G &= \frac{l!}{n!}(-1)^{l-n}\frac{r^{l-n-1}}{1-r} \\ &\quad + \frac{l!}{n!}\frac{(1-r)^{l-n-1}}{r} + l!\,l!\,(l-n)!\,(-1)^{l-n} \\ &\quad \times \sum_{i=1}^{l-n-1} \frac{(-1)^i r^{l-n-i-1}(1-r)^{i-1}}{(l-n-i)!\,i!\,(l-i)!\,(n+i)!}. \end{aligned} \tag{15}$$

The integrand in (13) is (15) multiplied by $D^{l-n}G$, but the series indicated in (15) gives a zero contribution to (13) because of the fact that

$$\int_0^1 r^\alpha\, D^\gamma G\, dr = 0 \tag{16}$$

whenever α is a nonnegative integer, $\alpha \leqslant \gamma - 1$, γ is a positive integer, and $\gamma < l$. Equation (16) is applicable to the series in (15) when the factor $(1-r)^{l-1}$ occurring there is expanded as a polynomial in r according to the binomial theorem. The remaining two terms on the right side of (15) give equal contributions to (13) as may be seen by a change of integration variable $r \to 1-r$ in one of them. Hence we obtain

$$c_n^{-2} = \frac{n!}{l!}\int_0^1 \frac{(1-r)^{l-n-1}}{r}\, D^{l-n}G\, dr. \tag{17}$$

Expansion of the factor $(1-r)^{l-n-1}$ in (17) via the binomial theorem and use of (16) reduces (17) to

$$c_n^{-2} = \frac{n!}{l!}\int_0^1 \frac{1}{r}\, D^{l-n}G\, dr. \tag{18}$$

The integral in (18) is easily evaluated and (7) obtained by applying in turn (11) and (14), and then an elementary identity on the resulting series of binomial coefficients. This completes the proof of (7).

We now give an explicit expression for the N-soliton solution, $u(x,t)$, of the KdV equation in the case of initial data given by (5). We accomplish this by evaluating the determinant, $M = \det[\mathscr{I} + C(x,t)]$, occurring in (4).

We introduce the notation $a_n \equiv c_n \exp(-nx + n^3 t) \equiv \mu_n^{-1/2}$, with c_n given by (7), and factor a_n out of row n and column n of M; this yields

$$M = (a_1 a_2 \cdots a_l)^2 \det[\,\mu_n \delta_{nm} + 1/(n+m)].$$

Expansion of this determinant gives

$$M = (a_1 a_2 \cdots a_l)^2 \sum_{\mathscr{J}} N_{\mathscr{J}}\, \mu_{\mathscr{J}}, \tag{19}$$

where the sum is over all possible sets (including the empty set $\emptyset$) of indices $\mathscr{J} = \{i_1, i_2, \ldots, i_k\}$ whose elements i_j are distinct positive integers with $i_j \leqslant l$ for each $j = 1,2,\ldots,k$, $k \leqslant l$. The symbol $\mu_{\mathscr{J}}$ denotes the product $\mu_{i_1}\mu_{i_2}\cdots\mu_{i_k}$, and $\mu_{\mathscr{J}} \equiv 1$ when $\mathscr{J} = \emptyset$. The symbol $N_{\mathscr{J}}$ denotes the k th-order minor obtained by deleting the rows and columns labelled by $i_1, i_2, \ldots i_k$ from the numerical determinant $|(n+m)^{-1}|$. In the case $k = l$, $N_{\mathscr{J}} \equiv 1$, and $N_\emptyset = \det[1/(n+m)] = c_1^{-2} c_2^{-2} \cdots c_l^{-2}$. Each $N_{\mathscr{J}}$ is a so-called double alternant determinant (a type studied by Cauchy[8] in 1841) and its value is given by a simple formula[9] for all l and $\mathscr{J}$:

$$N_{\mathscr{J}} = \frac{1}{2^{l-k} P_{\mathscr{J}_c}}\left(\frac{D_{\mathscr{J}_c}}{S_{\mathscr{J}_c}}\right)^2, \tag{20}$$

where $\mathscr{J}_c$ denotes the complement of $\mathscr{J}$ in $\{1,2,3,\ldots,l\}$, $P_{\mathscr{J}_c}$ is the product of all elements in $\mathscr{J}_c$, $D_{\mathscr{J}_c}$ is the product of all positive differences of elements in $\mathscr{J}_c$, and $S_{\mathscr{J}_c}$ is the product of all sums of elements in $\mathscr{J}_c$.

Hence the explicit form for $u(x,t)$ is

$$u(x,t) = -2(\ln M)_{xx} = -2(MM_{xx} - M_x^2)/M^2, \tag{21}$$

where M is given by (19) with the coefficients $N_{\mathscr{J}}$ of the sum in (19) given by (20).

In Appendix A we present the cases $l = 3$ and 4 in detail and illustrate the use of Eqs. (19) and (20).

The form of M expressed by (19) and (20) also applies to the general N-soliton case if $N_{\mathscr{J}}$ is understood to be obtained from the determinant $|1/(\kappa_n + \kappa_m)|$ and the sums, differences, and products in (20) are formed from the indicated sets of κ's. In that case the c_n's are no longer given by (7) and the μ_n's, $\mu_n^{-1/2} = c_n \exp(-\kappa_n x + \kappa_n^3 t)$, are also unspecified.

There are other simple reflectionless potentials described by Kundt and Newman and related to (2); they are all singular and for this reason were not considered by Kay and Moses. We describe them here for completeness and because they occur in some situations described in Sec. 4.

If we replace x by $x(-1)^{1/2}$ in (3), we get $u(x,0) = l(l+1)/\cosh^2 x$ and if we put, instead of c_n, $c_n(-1)^{1/2}$ in (3), we obtain

$$u(x,0) = \frac{l(l+1)}{\sinh^2 x}; \tag{22}$$

if we do both things the result is $u(x,0) = l(l+1)/\sin^2 x$. These are certainly reflectionless potentials and lead to *singular* N-soliton solutions. The inverse scattering method is not directly applicable to these cases because the potentials do not permit the usual scattering theory formalism to be introduced.

There is another simple reflectionless potential, $u(x,0) = l(l+1)/x^2$. In the next section it will be related to those discussed here.

3. THE WAVE EQUATION IN MINKOWSKI SPACE

"Reflectionless" potentials for the time-independent

Schrödinger equation are, of course, "nonscattering" potentials for the wave equation. The radial wave equation in characteristic coordinates,

$$\frac{\partial^2\psi}{\partial u\,\partial v} = \frac{l(l+1)}{(u+v)^2}\psi, \tag{23}$$

is certainly nonscattering. The coordinate transformation $u' = \tanh u$, $v' = \tanh v$ transforms (23) to

$$\frac{\partial^2\psi}{\partial u'\,\partial v'} = \frac{l(l+1)}{\sinh^2(u'+v')}\psi \tag{24}$$

while $u' = \tanh u$, $v' = \coth v$ transforms (23) to

$$\frac{\partial^2\psi}{\partial u'\,\partial v'} = \frac{l(l+1)}{\cosh^2(u'+v')}\psi. \tag{25}$$

Thus the reflectionless nature of $u(x,0) = -l(l+1)/\cosh^2 x$ and $u(x,0) = l(l+1)/\sinh^2 x$ is a simple consequence of that of $u(x,0) = l(l+1)/x^2$. If transformations involving the complex domain are permitted, then $l(l+1)/\cos^2 x$ and $l(l+1)/\sin^2 x$ are similarly related to $l(l+1)/x^2$. Kundt and Newman treated coordinate transformations among their nonscattering equations carefully, but seem to have overlooked these interconnections. Unfortunately we have been unable to find other such transformation which, if they existed, might significantly simplify the class of reflectionless potentials.

4. THE WAVE EQUATION IN SOME CURVED SPACES

The potential which produces the simplest N-soliton solution and its singular potential counterparts actually occur in a mathematically natural way in certain problems of interest in general relativity. These potentials occur when one considers the scalar wave equation on the background geometry due to a nonrotating black hole, in both the charged and uncharged cases. They also occur in the case of a rotating black hole; when axial symmetry is imposed on the scalar field this case is most like the others.

Consider the covariant scalar wave equation $g^{\mu\nu}\psi_{;\mu\nu} = 0$ where $g_{\mu\nu}$ is the Reissner–Nordström metric

$$ds^2 = \frac{\Delta}{r^2}dt^2 - \frac{r^2}{\Delta}dr^2 - r^2\,d\Omega^2, \tag{26}$$

$\Delta \equiv r^2 - 2mr + e^2$, (t,r,θ,ϕ) are the usual coordinates, and m and e are the mass and charge, respectively, of the black hole with $m > e \geqslant 0$. We make the usual separation of variables $\psi \to \psi(t,r)\,P_l^{m'}(\cos\theta)\,e^{im'\phi}$ and obtain the familiar radial wave equation

$$\frac{\Delta}{r^4}[\Delta\psi_r]_r - \psi_{tt} - \frac{\Delta}{r^4}l(l+1)\psi = 0. \tag{27}$$

Introduction of the variable

$$x \equiv \frac{1}{2}\ln\left|\frac{r-m-(m^2-e^2)^{1/2}}{r-m+(m^2-e^2)^{1/2}}\right| \tag{28}$$

eliminates the first derivative term from (27) and, in the region exterior to the horizon, $r > m + (m^2 - e^2)^{1/2}$, the resulting equation is

$$\psi_{xx} - \frac{r^4}{m^2-e^2}\psi_{tt} - \frac{l(l+1)}{\sinh^2 x}\psi = 0, \quad -\infty < x < 0, \tag{29}$$

with $r(x) = m - (m^2 - e^2)^{1/2}\coth x$. Inside the horizon, $m - (m^2 - e^2)^{1/2} < r < m + (m^2 - e^2)^{1/2}$, the equation is

$$\psi_{xx} - \frac{r^4}{m^2-e^2}\psi_{tt} + \frac{l(l+1)}{\cosh^2 x}\psi = 0, \quad -\infty < x < \infty, \tag{30}$$

with $r(x) = m - (m^2 - e^2)^{1/2}\tanh x$. The form of these equations in the Schwarzschild case is obtained by putting $e = 0$.

From the point of view of differential equations, (29) and (30) express an interesting property of scalar radiation which propagates in a black hole geometry, namely, that the radiation scattering caused by the curvature of spacetime is all contained in the one coefficient $r^4(x)$, known explicitly in terms of x, while the centrifugal term is one of the simple nonscattering potentials. Equation (29) has been used to calculate the transmission coefficient as a function of frequency for scalar radiation in a black hole geometry.[10,11]

Similar results and remarks apply to the case $m = e$ and the case when the background metric is Kerr with axial symmetry imposed on the field. When $m = e$ the radial wave equation in the exterior region, $r > m$, takes the form

$$\psi_{xx} - \frac{r^4}{m^2}\psi_{tt} - \frac{l(l+1)}{x^2}\psi = 0, \quad -\infty < x < 0, \tag{31}$$

with $r(x) = m(1 - 1/x)$. In the interior, $0 < r < m$, the same equations apply with the range of x being $1 < x < \infty$.

In the case of a Kerr background with spin angular momentum parameter a, the appropriate separated equation for a radial variable $R(r)$ is the Teukolsky[12] equation which, in the axially symmetric case, is

$$\Delta(\Delta R_r)_r + w^2\delta^2 R - l(l+1)\Delta R = 0, \tag{32}$$

where now $\Delta \equiv r^2 - 2mr + a^2$, and $\delta^2 \equiv (r^2 + a^2)^2 - a^2\Delta$. The separation leading to (32) is $\psi = e^{-iwt}e^{im'\phi}S(\theta)R(r)$, but $m' = 0$ because of axial symmetry. The coordinates (t,r,θ,ϕ) are Boyer–Lindquist coordinates.

For $m > a$, the variable x appropriate to (32) is that given by (28) with e replaced by a, and the resulting equation, in the exterior region $r > m + (m^2 - a^2)^{1/2}$, is

$$R_{xx} + \frac{w^2\delta^2}{m^2-a^2}R - \frac{l(l+1)}{\sinh^2 x}R = 0, \quad -\infty < x < 0. \tag{33}$$

In the interior region, a similar equation holds with the centrifugal term being $l(l+1)/\cosh^2 x$. When $m = a$ we have $x = (m - r)^{-1}$ and the centrifugal term is $-l(l+1)/x^2$.

APPENDIX A

We first calculate the KdV solution $u(x,t)$ from (19)–(21) in the case $l = 3$. The sum in (19) is, in this case,

$$\sum_{\mathscr{S}} N_{\mathscr{S}}\,\mu_{\mathscr{S}} = \mu_1\mu_2\mu_3 + N_{12}\mu_1\mu_2 + N_{13}\mu_1\mu_3 + N_{23}\mu_2\mu_3 + N_1\mu_1 + N_2\mu_2 + N_3\mu_3 + N_\emptyset, \tag{A1}$$

where, for example,

$$N_2 = \begin{vmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{6} \end{vmatrix},$$

which according to (20) has the value

$$N_2 = \frac{1}{2^2 \cdot 1 \cdot 3}\left(\frac{2}{4}\right)^2 = \frac{1}{48}.$$

Evaluating all coefficients we obtain from (19)

$$M = 1 + 6e^{2\eta_1} + 15e^{2\eta_2} + 10(e^{2\eta_3} + e^{2\eta_1 + 2\eta_2}) + 15e^{2\eta_1 + 2\eta_3} + 6e^{2\eta_2 + 2\eta_3} + e^{2\eta_1 + 2\eta_2 + 2\eta_3}, \quad \text{(A2)}$$

where $\eta_n \equiv -nx + n^3 t$.

We remove a factor of $e^{2\eta_1 + 2\eta_2 + 2\eta_3}$ from the numerator and denominator of (21) and then obtain from (21) and (A2) the solution

$$u(x,t) = -(T_3/24)\,[\cosh(\eta_1 + \eta_2 + \eta_3) + 6\cosh(\eta_2 + \eta_3 - \eta_1) + 15\cosh(\eta_1 + \eta_3 - \eta_2) + 10\cosh(\eta_1 + \eta_2 - \eta_3)]^{-2}, \quad \text{(A3)}$$

where

$$T_3 = 252 + \cosh 2(\eta_2 + \eta_3) + 10\cosh 2(\eta_1 + \eta_3) + 15\cosh 2(\eta_1 + \eta_2) + 30\cosh 2\eta_3 + 80\cosh 2\eta_2 + 50\cosh 2\eta_1 + 40\cosh 2(\eta_3 - \eta_1) + 25\cosh 2(\eta_3 - \eta_2) + 135\cosh 2(\eta_2 - \eta_1).$$

When $l = 4$ we find from (19)

$$M = 1 + 10e_1 + 45e_2 + (70e_3 + 50e_{12}) + (175e_{13} + 35e_4) + 126(e_{14} + e_{23}) + (175e_{24} + 35e_{123}) + (70e_{124} + 50e_{34}) + 45e_{134} + 10e_{234} + e_{1234}, \quad \text{(A4)}$$

where, for example, the notation is $e_{13} \equiv e^{2\eta_1 + 2\eta_3}$. Note that in (A4) the sum of coefficients of terms having the same x dependence of e^{-px} is a binomial coefficient $\binom{10}{p}$. This corresponds to the fact that for general l we have at $t = 0$

$$M(x,0) = (1 + e^{-2x})^{[1/2]l(l+1)}. \quad \text{(A5)}$$

Equation (A5) is easily obtained by integration of (5).

By direct calculation in (21), using (A4), an expression similar to (A3) for the solution $u(x,t)$ in the $l = 4$ case may be straightforwardly obtained. Because of its length, we omit giving this expression.

[1]R. Hirota, Phys. Rev. Lett. **27**, 1192 (1971).
[2]M. Wadati and M. Toda, J. Phys. Soc. Jpn. **32**, 1403 (1972).
[3]For a review see, for example, R.M. Miura, SIAM Rev. **18**, 412 (1976).
[4]I. Kay and H.E. Moses, J. Appl. Phys. **27**, 1503 (1956).
[5]W. Kundt and E.T. Newman, J. Math. Phys. **9**, 2193 (1968).
[6]See Ref. 3.
[7]L.D. Landau and E.M. Lifshitz, *Quantum Mechanics Nonrelativistic Theory*, 3rd ed. (Pergamon, New York, 1977), pp. 73–74.
[8]Thomas Muir, *The Theory of Determinants* (Dover, New York, 1960), Vol. 1, pp. 342–345.
[9]Thomas Muir, *The Theory of Determinants*, (Dover, New York, 1960), Vol. 3, p. 139. See also Ref. 2, Eq. (3.12).
[10]W.E. Couch and R.J. Torrence, Gen. Rel. Grav. **9**, 1119 (1978).
[11]W.E. Couch and R.J. Torrence, Gen. Rel. Grav. (to be published).
[12]S.A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).

Restoration of a functional from its functional derivatives

W. Garczyński and J. Hańćkowiak
Institute of Theoretical Physics, University of Wrocław, Cybulskiego 36, 50-205 Wrocław, Poland

(Received 5 April 1978)

Inverse operation to a functional differentiation is considered.

INTRODUCTION

In this paper we dwell on the problem of finding a functional, knowing its functional derivatives. It resembles the problem of finding a primitive function from its derivative, i.e., an integral. Obviously, it is not a unique procedure since corresponding "constants of integration," enter the solution. Problems of this sort arise in various branches of physics using functional methods, and in the quantum field theory as well.

The paper consists of two parts. The first part deals with a general case of a functional being just a differentiable. The second part deals with the so called analytic functionals or element of the Bargmann space.

1. RESTORATION OF A FUNCTIONAL FROM ITS FUNCTIONAL DERIVATIVES-A GENERAL APPROACH

Let $F[q]$ be a complex-valued functional on space of functions $q(x)\in\mathcal{F}$, $x\in R^4$,

$$F:\mathcal{F}\to\mathbb{C}. \tag{1.1}$$

One may define an associate function $f(z)$ of a complex variable z as follows:

$$f(z)=F[zq],\quad q \text{ fixed}. \tag{1.2}$$

Definition 1: A functional F is called an analytic in a region D if the function $f(z)$ is analytic in D for any $q\in\mathcal{F}$.

Definition 2: A functional F is called regular in D if this region is a simply connected set and $f(z)$ is regular there for any $q\in\mathcal{F}$.

In the last case we may write for $z,z_0\in D$ the Cauchy formula

$$f(z)=\frac{1}{2\pi i}\oint_{\partial\mathcal{D}}\frac{f(z')}{z'-z}dz'=\sum_{n=0}^{\infty}a_n(z-z_0)^n, \tag{1.3}$$

where

$$a_n=\frac{f^{(n)}(z_0)}{n!}=\frac{1}{2\pi i}\oint_{\partial\mathcal{D}}\frac{f(z')}{(z'-z_0)^{n+1}}dz'=a_n[q,z_0]. \tag{1.4}$$

Putting $z=1$ one gets the representation for $F[q]$

$$F[q]=f(1)=\sum_{n=0}^{\infty}a_n[q,z_0](1-z_0)^n, \tag{1.5}$$

where the coefficients are given in terms of functional derivatives of F.

$$a_0=[q,z_0]=f(z_0)=F[z_0q]$$

$$a_1[q,z_0]=\frac{f^{(1)}(z_0)}{1!}=\int d^4x\,q(x)\frac{\delta F[p]}{\delta p(x)}\bigg|_{p=z_0q}, \tag{1.6}$$

...

$$a_n[q,z_0]=\frac{f^{(n)}(z_0)}{n!}$$

$$=\frac{1}{n!}\int d^4x_1\cdots d^4x_n\,q(x_1)\cdots q(x_n)$$

$$\times\frac{\delta^n F[q]}{\delta p(x_1)\cdots\delta p(x_n)}\bigg|_{p=z_0q}$$

$$=\frac{1}{n!}\left(q\cdot\frac{\delta}{\delta p}\right)^n F[p]\bigg|_{p=z_0q}.$$

We used here the obvious notation

$$f\cdot g=\int d^4x f(x)g(x). \tag{1.7}$$

In particular, when $z_0=0$ we get the Volterra expansion

$$F[q]=\sum_{n=0}^{\infty}a_n[q,0]$$

$$=\sum_{n=0}^{\infty}\frac{1}{n!}\left(q\cdot\frac{\delta}{\delta p}\right)^n F[p]\bigg|_{p=0}. \tag{1.8}$$

The nth coefficient $a_n[q,0]$ is the nth order monomial functional in variable q.

Now, let the function $f(z)$ be regular in a ring formed by two concentric circles C_{R_1}, C_{R_2} of radiuses R_1, R_2, respectively ($R_1<R_2$), with common center at z_0. Then $f(z)$ may be expanded into the Laurent series

$$f(z)=\sum_{n=-\infty}^{\infty}a_n(z-z_0)^n, \tag{1.9}$$

where

$$a_n=\frac{1}{2\pi i}\oint_{C_K}\frac{f(z')dz'}{(z'-z_0)^{n+1}}=a_n[q,z_0], \tag{1.10}$$

$$a_{-n}=\frac{1}{2\pi i}\oint_{C_K}f(z')(z'-z_0)^{n-1}dz'=a_{-n}[q,z_0]. \tag{1.11}$$

Putting $z=1$ we obtain for the functional $F[q]$ the expansion

$$F[q]=\sum_{n=-\infty}^{\infty}a_n[q,z_0](1-z_0)^n. \tag{1.12}$$

 0022-2488/79/122427-04$01.00

It attains a simpler form when $z_0 = 0$. Obviously, one assumes tacitly that the point $z = 1$ lies in the analyticity domain of $f(z)$. Terms with negative n give those parts of the functional which may be pictured as inverses of monomial functionals represented by a_n with positive index.

Let us consider now a more general case when a functional $F[q]$ might be a less regular one.

Definition 3: A functional $F[q]$ is a smooth one of degree n when the function

$$f(\xi) = F[\xi q], \quad \xi \in R, \tag{1.13}$$

is the n-times differentiable function in the vicinity of the point $\xi = 0$.

In such a case one may use the Taylor expansion

$$f(1) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} + R_n(1), \tag{1.14}$$

where the remaining term $R_n(1)$ is given by the formula[1]

$$R_n(1) = \frac{1}{n!} \int_0^1 d\lambda\, (1-\lambda)^n f^{(n+1)}(\lambda). \tag{1.15}$$

Here, in fact, one needs also to assume the existence of the $f^{(n+1)}$ derivative.

On account of the equalities

$$f(1) = F[q],$$

$$f^{(1)}(0) = \left(q \cdot \frac{\delta}{\delta p}\right) F[p]\Big|_{p=0}, \tag{1.16}$$

...

$$f^{(n)}(0) = \left(q \cdot \frac{\delta}{\delta p}\right)^n F[p]\Big|_{p=0}$$

we obtain the following formula:

$$F[q] = \sum_{k=0}^{n} \frac{1}{k!} \left(q \cdot \frac{\delta}{\delta p}\right)^k F[p]\Big|_{p=0} + \frac{1}{n!} \int_0^1 d\lambda\, (1-\lambda)^n \left(q \cdot \frac{\delta}{\delta p}\right)^{n+1} F[p]\Big|_{p=\lambda q}. \tag{1.17}$$

This formula permits us to restore a functional $F[q]$ when the derivatives

$$\left(q \cdot \frac{\delta}{\delta p}\right)^k F[p]\Big|_{p=0}, \quad k = 0,1,\ldots,n,$$
$$\left(q \cdot \frac{\delta}{\delta p}\right)^{n+1} F[p]\Big|_{p\neq 0} \tag{1.18}$$

are known. In particular, if $F[0]$ and the first derivative $\delta F[q]/\delta q(x) = F'[q;x]$ are given we may restore the functional $F[q]$ as follows:

$$F[q] = F[0] + \int_0^1 d\lambda\, (q \cdot F'[\lambda q]),$$
$$F'[q;x] = \frac{\delta F[q]}{\delta q(x)}. \tag{1.19}$$

A consistency condition must be fulfilled, of course,

$$F''[q;x,y] = F''[q;y,x]. \tag{1.20}$$

For instance, a generating functional $\tau[p]$ of τ-functions, of a real scalar field, satisfies the Schwinger equation[2]

$$\tau'[p;x] = i \int d^4y\, D^c(x-y) \times \left\{p(y) + L'_{\text{int}}\left[-i\frac{\delta}{\delta p};y\right]\right\}\tau[p] \tag{1.21}$$

as the result of the formula[3]

$$\tau[p] = N^{-1} \exp\left(iL_{\text{int}}\left[-i\frac{\delta}{\delta p}\right]\right)\exp\left(\frac{i}{2} p \cdot D^c p\right). \tag{1.22}$$

The normalization factor N is determined by the condition

$$\tau[0] = 1. \tag{1.23}$$

The interaction functional $L_{\text{int}}[p]$ is connected with the interaction Lagrangian $L_{\text{int}}[p(x)]$ as follows,

$$L_{\text{int}}[p] = \int d^4x\, L_{\text{int}}[p(x)] \tag{1.24}$$

while the causal function $D^c(x)$ is given by the formula

$$D^c(x) = (2\pi)^{-4} \int d^4p (m^2 - p^2 - i\epsilon)^{-1} \exp(-ipx). \tag{1.25}$$

Finally, the symbol $L'_{\text{int}}[-i(\delta/\delta p);y]$ means

$$L'_{\text{int}}\left[-i\frac{\delta}{\delta p};y\right] = \frac{\delta L_{\text{int}}[q]}{\delta q(y)}\Big|_{q=-i(\delta/\delta p)}. \tag{1.26}$$

The consistency condition (1.20) reads in this case

$$\left\{L''_{\text{int}}\left[-i\frac{\delta}{\delta p};x,y\right] - L''_{\text{int}}\left[-i\frac{\delta}{\delta p};y,x\right]\right\}\tau[p] = 0. \tag{1.27}$$

It is always fulfilled if an interaction functional is a regular one

$$L_{\text{int}}[p] = \sum_{n=0}^{\infty} l_n \int d^4x\, p^n(x). \tag{1.28}$$

Taking into account the formula (1.23) we get from (1.19) an integral form of the Schwinger equation

$$\tau[p] = 1 + ip \cdot D^c \cdot p \int_0^1 d\lambda\, \lambda\tau[\lambda p] + i\int_0^1 d\lambda\, p \cdot D^c \cdot L'_{\text{int}}\left[-i\frac{\delta}{\delta \lambda p}\right]\tau[\lambda p]. \tag{1.29}$$

This equation may be a starting point for developping an approximate method for the evaluation of the generating functional $\tau[p]$.[4]

2. RESTORATION OF A FUNCTIONAL FROM ITS FUNCTIONAL DERIVATIVES CONSIDERED AS OPERATORS IN THE BARGMANN SPACE

A functional derivative may be consider as an operator in the Bargmann space of functional power series.[5] Namely,

any element $V[\alpha]$ of this space is a regular functional of the form

$$V[\alpha] = \sum_{n=0}^{\infty} \frac{1}{n!} \int d^4x_1 \cdots d^4x_n v_n(x_1,\ldots,x_n)\alpha(x_1)\cdots\alpha(x_n)$$

$$\equiv \sum_{n=0}^{\infty} \frac{1}{n!} \int dx_n v_n(x_n) e[\alpha;x_n], \tag{2.1}$$

$\alpha(x)$ is some function from $\mathcal{F}$. (We use the Rzewuski notation adopted in Ref. 3.) Its norm is defined by the scalar product in the following way:

$$V^*V = \|V\|^2 = V\left[\frac{\delta}{\delta p}\right] V[p]\Big|_{p=0}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int d^4x_n\, v_n^2(x_n) < \infty. \tag{2.2}$$

Any operator A in this space has the form of a functional matrix

$$A \leftrightarrow A[\alpha,\beta] \tag{2.3}$$

$$A[\alpha,\beta] = \sum_{m,n=0}^{\infty} \frac{1}{\sqrt{m!n!}} \int dx_m\, dy_n\, a_{mn}(x_m,y_n)$$

$$\times e[\alpha;x_m] e[\beta;y_n]. \tag{2.4}$$

The action of A on vector V yields another vector V' according to the rule

$$V'[\alpha] = (A^*V)[\alpha] = A\left[\alpha, \frac{\delta}{\delta p}\right] V[p]\Big|_{p=0}. \tag{2.5}$$

A unit matrix for this multiplication law is given by the formula

$$I[\alpha,\beta] = e^{\alpha\cdot\beta}. \tag{2.6}$$

A matrix representing functional differentiation is given by

$$D[x;\alpha,\beta] = \beta(x) e^{\alpha\cdot\beta}, \tag{2.7}$$

since we have

$$(D[x]^*V)[\alpha]$$

$$= D\left[x;\alpha,\frac{\delta}{\delta p}\right] V[p]\Big|_{p=0} = \frac{\delta V[\alpha]}{\delta\alpha(x)}. \tag{2.8}$$

Therefore, finding $V[\alpha]$ from this equality is equivalent to finding an inverse operation (inverse matrix) to $D[x]$. More precisely, we are looking for a matrix $D^{-1}[x;\alpha,\beta]$ satisfying the relation

$$\int d^4x\, D^{-1}[x]^*D[x] = 1 - P_0, \tag{2.9}$$

where P_0 is a projection onto subspace of scalar, which coincides with a kernel of the matrix $D[x]$

$$P_0[\alpha,\beta] = 1, \quad P_0^*P_0 = P_0. \tag{2.10}$$

Going over to the matrix elements we obtain from (2.9) the equation

$$\int d^4x\, D^{-1}\left[x;\alpha,\frac{\delta}{\delta p}\right] D[x;p,\beta]\Big|_{p=0} = e^{\alpha\cdot\beta} - 1. \tag{2.11}$$

One easily finds that the following matrix solves it:

$$D^{-1}[x;\alpha,\beta] = \alpha(x)\frac{e^{\alpha\cdot\beta}-1}{\alpha\cdot\beta} \equiv \alpha(x)E_1(\alpha\cdot\beta), \tag{2.12}$$

where we denoted the standard function

$$E_1(x) = (e^x - 1)/x. \tag{2.13}$$

This result and its generalization to higher order derivatives follows from the formulas

$$V[\alpha] = (I^*V)[\alpha] = e^{\alpha(\delta/\delta p)} V[p]\,|_{p=0}$$

$$= \left[e^{\alpha(\delta/\delta p)}1 - \frac{1}{1!}\left(\alpha\cdot\frac{\delta}{\delta p}\right) - \cdots\right.$$

$$\left. - \frac{1}{(n-1)!}\left(\alpha\cdot\frac{\delta}{\delta p}\right)^{n-1}\right] V[p]\Big|_{p=0}$$

$$+ \left[1 + \frac{1}{1!}\left(\alpha\cdot\frac{\delta}{\delta p}\right) + \cdots\right.$$

$$\left. + \frac{1}{(n-1)!}\left(\alpha\cdot\frac{\delta}{\delta p}\right)^{n-1}\right] V[p]\Big|_{p=0}$$

$$= E_n\left(\alpha\frac{\delta}{\delta p}\right)\left(\alpha\cdot\frac{\delta}{\delta p}\right)^n V[p]\Big|_{p=0}$$

$$+ \sum_{k=0}^{n-1} \frac{1}{k!}\left(\alpha\cdot\frac{\delta}{\delta p}\right)^k V[p]\Big|_{p=0}, \tag{2.14}$$

where again we denoted the standard function

$$E_n = \frac{e^x - 1 - x/1! - \cdots - x^{n-1}/(n-1)!}{x^n}. \tag{2.15}$$

Thus, knowing $V[0]$, $V^{(1)}[0,x_1],\ldots,V^{(n-1)}[0,x_1,\ldots,x_n]$ and $V^{(n)}[\alpha,x_1,\ldots,x_n]$ we may restore the whole functional $V[\alpha]$.

A functional derivative of nth order is represented by the matrix

$$D[x_1,\ldots,x_n;\alpha,\beta] = \beta(x_1)\cdots\beta(x_n)e^{\alpha\cdot\beta}, \tag{2.16}$$

$$(D[x_1,\ldots,x_n]^*V)[\alpha] = \frac{\delta^n V[\alpha]}{\delta\alpha(x_1)\cdots\delta\alpha(x_n)}, \tag{2.17}$$

and an inverse operation is represented by

$$D^{-1}[x_1,\ldots,x_n;\alpha,\beta] = \alpha(x_1)\cdots\alpha(x_n)E_n(\alpha\cdot\beta). \tag{2.18}$$

The following equality holds,

$$\int d^4x_1\cdots d^4x_n\, D^{-1}[x_1,\ldots,x_n]^*D[x_1,\ldots,x_n]$$

$$= I - \sum_{k=0}^{n-1} P_k, \tag{2.19}$$

where P_k is a projection onto a subspace of k th order monomial functionals

$$P_k[\alpha,\beta] = (1/k!)(\alpha\cdot\beta)^k, \quad k = 0,1,\ldots,n-1. \tag{2.20}$$

one easily checks that

$$P_k * P_j = \delta_{kj} P_k, \tag{2.21}$$

$$P_k\left[\alpha, \frac{\delta}{\delta p}\right] p(x_1)\cdots p(x_j)\Big|_{p=0} = \delta_{kj}\alpha(x_1)\cdots\alpha(x_k) \tag{2.22}$$

and the completeness condition holds

$$\sum_{k=0}^{\infty} P_k = I. \tag{2.23}$$

The above considerations may be somewhat generalized for, so called, projected derivatives.[4] Namely, if π is a projection operator in the function space $\mathcal{F}$

$$(\pi\alpha)(x) = \int d^4y\, \pi(x,y)\alpha(y) = (\pi^2\alpha)(x), \tag{2.24}$$

where the kernel $\pi(x,y)$ satisfies the composition rule

$$\int d^4z\, \pi(x,z)\pi(z,y) = \pi(x,y) \tag{2.25}$$

then, we define the projected derivative as follows

$$\left(\pi\cdot\frac{\delta}{\delta\alpha}\right)(x) = \int d^4y\, \pi(x,y)\frac{\delta}{\delta\alpha(y)}, \tag{2.26}$$

The corresponding functional matrix is

$$D_\pi[x;\alpha,\beta] = (\pi\beta)(x)e^{\alpha\cdot\beta}. \tag{2.27}$$

Its null space is now richer than before since the equation

$$\left(\pi\cdot\frac{\delta}{\delta\alpha}\right)(x)N[\alpha] = \int d^4y\, \pi(x,y)N'[\alpha;y] = 0 \tag{2.28}$$

is satisfied not only by scalars.

To describe this space we introduce a complementing projection operator σ such as

$$\sigma = 1 - \pi, \tag{2.29}$$

$$\sigma(x,y) = \delta(x-y) - \pi(x,y). \tag{2.30}$$

One verifies easily that

$$\sigma^2 = \sigma = \sigma^T, \quad \sigma\pi = 0 = \pi\cdot\sigma,$$

$$\int dz\, \sigma(x,z)\pi(z,y) = 0. \tag{2.31}$$

Therefore, any vector of the form

$$N[\alpha] = e^{\alpha\cdot\sigma\cdot(\delta/\delta p)}M[p]\big|_{p=0} = (P_\sigma * M)[\alpha] \tag{2.32}$$

satisfies Eq. (2.28) for arbitrary $M[p]$. When the operator σ vanishes, i.e., when $\pi = 1$, we get for the solutions of (2.28) the scalars, $M = M[p]|_{p=0}$ as it should be.

Here P_σ stands for a projection operator onto the null space of the projected derivative

$$P_\sigma[\alpha,\beta] = e^{\alpha\cdot\sigma\cdot\beta}, \quad P_\sigma * P_\sigma = P_\sigma. \tag{2.33}$$

The inverse of a projected derivative is represented by the matrix

$$\begin{aligned} D_\pi^{-1}[x;\alpha,\beta] &= \alpha(x)\frac{e^{\alpha\cdot\beta} - e^{\alpha\sigma\beta}}{\alpha\pi\beta} \\ &= \alpha(x)\frac{e^{\alpha\pi\beta} - 1}{\alpha\pi\beta}e^{\alpha\sigma\beta} \\ &= \alpha(x)E_1(\alpha\pi\beta)P_\sigma[\alpha,\beta]. \end{aligned} \tag{2.34}$$

It satisfies the condition

$$\begin{aligned} &\int d^4x\, D_\pi^{-1}\left[x;\alpha,\frac{\delta}{\delta p}\right]D_\pi[x;p,\beta]\Big|_{p=0} \\ &\quad = e^{\alpha\beta} - e^{\alpha\cdot\sigma\cdot\beta} \end{aligned} \tag{2.35}$$

or in a matrix form

$$\int d^4x D_\pi^{-1}[x] * D_\pi[x] = I - P_\sigma. \tag{2.36}$$

Therefore, if we have to find a functional $V[\alpha]$ knowing its first projected derivative

$$\int d^4y\, \pi(x,y)\frac{\delta V[\alpha]}{\delta\alpha(y)} = D_\pi[x] * V[\alpha] \tag{2.37}$$

we apply to both sides the operation $D_\pi^{-1}[x]$ and use the relation (2.35). We get as a result

$$\begin{aligned} &V[\alpha] - V[\sigma\alpha] \\ &\quad = \int d^4x\, d^4y\, D_\pi^{-1}\left[x;\alpha,\frac{\delta}{\delta p}\right]\pi(x,y)V'[p,y]\Big|_{p=0}, \end{aligned} \tag{2.38}$$

$$D_\pi^{-1}\left[x;\alpha,\frac{\delta}{\delta p}\right] = \alpha(x)\frac{e^{\alpha\pi(\delta/\delta p)} - e^{\alpha\sigma(\delta/\delta p)}}{\alpha\pi(\delta/\delta p)}e^{\alpha\sigma(\delta/\delta p)}. \tag{2.39}$$

In a subsequent publication we will describe some applications of techniques presented here.

[1] I.C. Gradstein and I.M. Ryzik, *Tablitsy integralov, sum. rjadov i proizvedenij* (Nauka, Moscow, 1971), Chap. 5.
[2] J. Schwinger, Proc. Natl. Acad. Sci. USA **37**, 452 (1951).
[3] J. Rzewuski, *Field Theory II* (Polish Scientific Publisher, Warsaw, 1969 and Iliffe Books).
[4] J. Hańćkowiak, preprints, Wrocław University (1978).
[5] V. Bargmann, Comm. Pure Appl. Math. **14**, 187 (1961).

Visual geometry and the algebraic properties of spinors

H. Hellsten
Institute of Theoretical Physics, University of Stockholm, Vanadisvägen 9, S-113 46 Stockholm, Sweden

(Received 26 April 1977; accepted for publication 12 March 1979)

Key words of the present paper are visualization and concrete geometry. We develop ordinary two-component spinor algebra from a concrete geometrical model of spinor space. The null flag picture may be said to constitute such a model as far as topological and differential properties of spinors go. In our model it is also possible to visualize the *algebraic* properties of spinors by straightforward geometrical constructions. Notably, an interpretation of spinor addition is given in terms of a geometrical procedure, analogous to the addition of real 3-vectors via the parallelogram rule. By this procedure the relation between the projection of spinors on the 2-sphere and the projection of their sum can directly be read off. The connection between null flags and our presentation of spinors is touched upon. It is planned to discuss the connection to Minkowski space more closely in a forthcoming paper.

1. INTRODUCTION

The present paper has arisen from a critical attitude towards the "formalistic method" in the foundations of physics. In opposition to such a method we hold that the primitive terms of a physical theory should be of a material or visual nature. [1] The necessity of invoking spinor algebra in the physical description of nature could perhaps be regarded as an argument favoring the purely formalistic view, since spinors are normally not believed to bear very directly on material objects. As we shall make evident, such an abstract inclination towards spinors is quite unnecessary, and it is possible to develop spinor algebra in an entirely visual way, analogous to the material interpretation of the algebra of Euclidean 3-vectors.

The search for alternative ways of describing spinors seems to have a history as old as spinor algebra itself (cf. the works of Cartan, [2] Veblen, [3] Penrose, [4] and others). Notably, in the view adopted by Penrose, spinors are represented as null flags, the flag pole being a null vector and the flag itself tangent to the null cone. However, even though this is an extremely useful way of interpreting spinors, we do not regard it to be the final word concerning their visual content. Firstly, this interpretation has the fault of leaving spinors visually undetermined within a sign. A well-known remedy is to interpret spinors as elements in the universal covering of the space of null flags. In this way we obtain a *topologically* faithful interpretation of spinors. Still we lack visual insight in their *algebraic* properties. For example, it would be very awkward to try to form the sum of two spinors, or to resolve a spinor into base components, resting the argument on null flags. [5] The actual structure of spinor space as a linear, complex 2-space $\mathbf{C}^2$, is not at all evident from this interpretation. In contrast, the visualization of Euclidean 3-vectors as directed arrows truly reflects *all* the algebraic properties of Euclidean 3-space (addition is visualized by the parallelogram rule, etc.). To obtain an equally faithful interpretation of spinors we shall radically divert from the null flag interpretation. In the approach to spinor algebra presented in this paper, spinors are identified with certain constructions carried out in Euclidean 3-space. The notable feature about these constructions is that they are defined in such a way that the computational rules for Euclidean 3-space will induce all the computational rules for spinors. In carrying out this program, Euclidean 3-space will be regarded as a formal structure. Thus the spinor picture to be presented is very much a *logical* reformulation of spinor algebra into the formal language of Euclidean 3-geometry. However, since we already know how to visualize the formal properties of Euclidean 3-space, the visual meaning of spinor algebra will become clear when given this reformulation.

2. THE SPINOR SPACE MODEL

Regard a 2-plane $\mathbf{C}_P$ possessing a preferred vector $1_P \in \mathbf{C}_P$ of unit length. To us $\mathbf{C}_P$ will serve as a model of a *linear, complex 1-space*. Thus the complex numbers $\mathbf{C}$ will act as scalars, an element $\rho e^{i\alpha} \in \mathbf{C}$ causing in $\mathbf{C}_P$ a rotation through the angle α and a rescaling by the real factor ρ. Denote an arbitrary element in $\mathbf{C}_P$ by z_P and scalar action upon this element by $z' z_P$ (where $z' \in \mathbf{C}$). We also adopt exponential notation in $\mathbf{C}_P$, defining $e_P^{i\alpha} \equiv e^{i\alpha}\, 1_P$. Our spinor model will be based upon a fixed 2-sphere S^2 of unit diameter, imbedded in Euclidean 3-space E^3 (cf. Fig. 1). An arbitrary but fixed point P is selected on the 2-sphere, and the following three definitions made with respect to this point:

The plane $\mathbf{C}_P$: We now restrict this symbol to stand for the tangent plane of S^2 at P, containing the preferred unit vector $1_P \in \mathbf{C}_P$ and given the structure of a linear, complex 1-space in the way indicated.

The half-space W_P: This symbol will denote the topologically closed half-space, which contains $\mathbf{C}_P$ as its boundary and includes S^2. Note that this half-space is algebraically closed under vector addition and scalar multiplication by elements of $\mathbf{R}_+$ (the real numbers $\geqslant 0$).

The bilinear product $W_P \otimes \mathbf{C}_P$ It may be checked that the following three axioms, defining a rather special form of tensor product, are mutually consistent:

(A) $\rho_1 (x_P \otimes z_P) + \rho_2 (y_P \otimes z_P) = (\rho_1 x_P + \rho_2 y_P) \otimes z_P$,

 0022-2488/79/122431-07$01.00

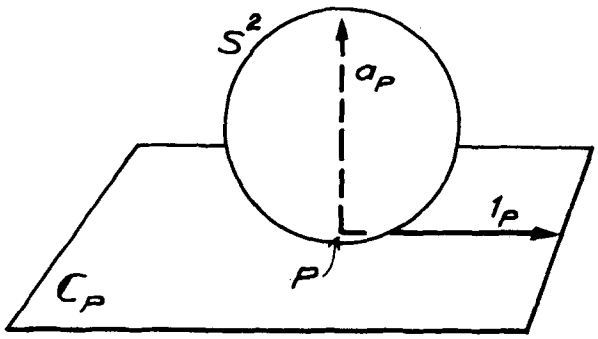

FIG 1. Basic definitions of $W_P \otimes \mathbf{C}_P$.

(B) $z_1(x_P \otimes z_P) + z_2(x_P \otimes z'_P) = x_P \otimes (z_1 z_P + z_2 z'_P)$,

(C) $z_P \otimes (zz'_P) = (\bar{z}z_P) \otimes z'_P$.

Here $x_P, y_P \in W_P$; $z_P, z'_P \in \mathbf{C}_P$; $\rho_1, \rho_2 \in \mathbf{R}_+$, and $z, z_1, z_2 \in \mathbf{C}$. Bar denotes complex conjugation.

From axioms A and B it follows that $x_P \otimes \rho z_P = (\rho x_P) \otimes z_P$ for $\rho \in \mathbf{R}_+$. Thus the expression $x_P \otimes z_P$ may be geometrically interpreted as a vector pair $\rho x_P, z_P/\rho$ in Euclidean 3-space, determined up to an arbitrary choice of the positive scale factor ρ. Axiom C completes A and B in stating that in the degenerate case when both factors of $x_P \otimes z_P$ lie in $\mathbf{C}_P$, we may not only change their relative length, but we are also free to rotate x_P and z_P in $\mathbf{C}_P$, keeping the angle between them constant.

The elements of $W_P \otimes \mathbf{C}_P$ are to be regarded as spinors, defined in the special terms of visual geometry. This will be justified in the next section, where we shall show $W_P \otimes \mathbf{C}_P$ to possess the two important properties of a spinor space. Firstly, it will be demonstrated how the rules A, B, C turn $W_P \otimes \mathbf{C}_P$ into a linear, complex 2-space. Secondly, a mapping of $W_P \otimes \mathbf{C}_P$ (or more precisely, the non-zero elements of $W_P \otimes \mathbf{C}_P$) on S^2 will be defined. Algebraically, this mapping is the standard projection of a linear, complex 2-space on a 2-sphere. It may also be obtained in any of the well-known ways, e.g., from the conformal plane of the 2-sphere or from the connection between spinors and null vectors of Minkowski space. It has the property of inducing ordinary rotations of the 2-sphere when unitary transformations act upon the 2-space.

3. SPINOR SPACE STRUCTURE OF $W_P \otimes \mathbf{C}_P$

Our first step is to prove that $W_P \otimes \mathbf{C}_P$ and $\mathbf{C}^2$ are canonically isomorphic. The reader is again referred to Fig. 1. As is seen in this figure the unit normal of $\mathbf{C}_P$, pointing into W_P, is denoted a_P. An arbitrary element of W_P may be decomposed into a vector along a_P and a vector in $\mathbf{C}_P$, the element thus written $\rho a_P + z1_P$ for some unique pair of numbers $\rho \in \mathbf{R}$, $z \in \mathbf{C}$. In a corresponding manner we write for an arbitrary element of $W_P \otimes \mathbf{C}_P$ [making use of the rules (A), (C) in turn]

$$(\rho a_P + z1_P) \otimes z'1_P = (\rho a_P) \otimes z'1_P + (z1_P) \otimes z'1_P = a_P \otimes (\rho z')1_P + 1_P \otimes (\bar{z}z')1_P. \quad (3.1)$$

Proceeding inversely, we start with an arbitrary pair of complex numbers $z_1, z_2 \in \mathbf{C}$ to which we may assign a unique element of $W_P \otimes \mathbf{C}_P$ as follows:

$$a_P \otimes z_1 1_P + 1_P \otimes z_2 1_P = a_P \otimes z_1 1_P + (\bar{z}_2 \bar{z}_1^{-1} 1_P) \otimes z_1 1_P = (a_P + \bar{z}_2 \bar{z}_1^{-1} 1_P) \otimes z_1 1_P \quad (3.1a)$$

(here $z_1 \neq 0$; in the case $z_1 = 0$, the corresponding element of $W_P \otimes \mathbf{C}_P$ is just $1_P \otimes z_2 1_P$). Evidently, (3.1) constitutes a *unique* decomposition of the elements of $W_P \otimes \mathbf{C}_P$. Defining

$$a_P \otimes z1_P \cong \begin{pmatrix} z \\ 0 \end{pmatrix}, \quad 1_P \otimes z1_P \cong \begin{pmatrix} 0 \\ z \end{pmatrix}, \quad (3.2)$$

we obtain a one-to-one mapping between $W_P \otimes \mathbf{C}_P$ and $\mathbf{C}^2$, transforming components into components. According to (B), the transformation of components is linear, thus the mapping is an *isomorphism*. It follows that any calculation, normally performed in $\mathbf{C}^2$, could just as well be carried out in $W_P \otimes \mathbf{C}_P$. For instance, spinor addition corresponds in $W_P \otimes \mathbf{C}_P$ to evaluation of

$$x_P \otimes z_P + y_P \otimes z'_P.$$

In order to carry out this operation, we may apply (3.1) to each term of the expression and then rule (B) and (3.1a) in turn. Alternatively, we can rely on the geometric character of $W_P \otimes \mathbf{C}_P$ in order to obtain a *component-independent* construction of the sum (analogous to the addition of Euclidean vectors according to the parallelogram rule). The full discussion of this procedure must wait until the developments of Sec. 6, however.

Apart from carrying the structure of a vector space, $W_P \otimes \mathbf{C}_P$ provides a natural way to project its nonzero elements on the 2-sphere S^2. It will be convenient to introduce affine vectors PQ, where Q is a point on S^2 not equal to P. Any element x_P in the *interior* of W_P may clearly be written $x_P = \rho PQ$ for some $Q \in S^2$ and some positive factor ρ. The corresponding expression $x_P \otimes z_P$ may be rewritten uniquely

$$x_P \otimes z_P = PQ \otimes \rho z_P. \quad (3.3)$$

We define Q to be the projection of (3.3) *on* S^2. Continuity in the limit $Q \to P$ demands P to be the projection on S^2 of all nonzero elements of the form $z_P \otimes z'_P$; $z_P, z'_P \in \mathbf{C}_P$. The zero element of $W_P \otimes \mathbf{C}_P$ may be written in any of the following ways:

$$PQ \otimes 0_P = z_P \otimes 0_P = 0_P \otimes 0_P$$

($Q \in S^2$, $z_P \in \mathbf{C}_P$) and thus cannot be invariantly related to any particular point on S^2. This completes the definition of the projection mapping. [6]

The conformal plane of Fig. 2 is the plane parallel to $\mathbf{C}_P$, which meets S^2 at the end point of the vector a_P (we may thus write its elements $a_P + z\,1_P$, $z \in \mathbf{C}$). It is related to S^2 by ordinary conformal projection. We may write for any nonzero $PQ \otimes z_P$,

$$PQ \otimes z_P = (a_P + z'1_P) \otimes \rho' z_P, \quad (3.4)$$

where $\rho' \in \mathbf{R}_+$ is some scale factor and $z' \in \mathbf{C}$ is related to $Q \in S^2$ as a complex number to the corresponding point of the Riemann sphere. Let us study how the conformal plane behaves under the action of a unitary transformation of $W_P \otimes \mathbf{C}_P$.

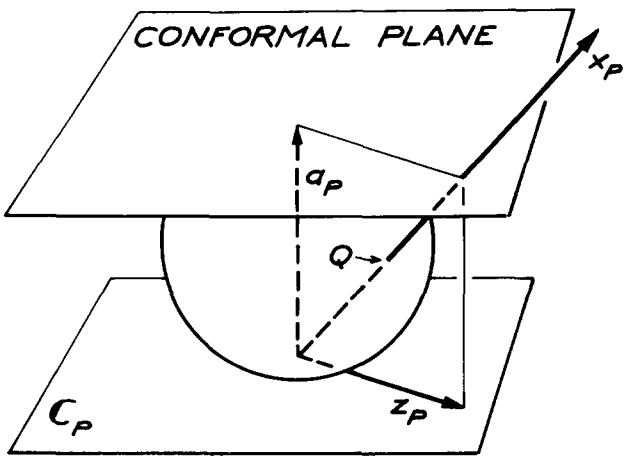

FIG.2. Relation betwen $\mathbf{C}_P$, S^2, and conformal plane.

The action of

$$\begin{pmatrix} \bar{\alpha} & \beta \\ -\bar{\beta} & \alpha \end{pmatrix} \tag{3.5}$$

($\alpha, \beta \in \mathbf{C}$) on $W_P \otimes \mathbf{C}_P$ is defined by (3.2). We obtain for non-zero elements $x_P \otimes z_P$, written in the form (3.4)

$$\begin{aligned}
&(a_P + z1_P) \otimes z'1_P \\
&\quad = z'(a_P \otimes 1_P) + \bar{z}z'(1_P \otimes 1_P) \\
&\quad \to z'[\bar{\alpha}(a_P \otimes 1_P) - \bar{\beta}(1_P \otimes 1_P)] \\
&\qquad + \bar{z}z'[\beta(a_P \otimes 1_P) \\
&\qquad + \alpha(1_P \otimes 1_P)] \\
&\quad = a_P \otimes (\bar{\alpha}z' + \beta z'\bar{z})1_P + 1_P(\bar{\alpha}z\bar{z}' - \beta\phi\bar{z}') \otimes 1_P \\
&\quad = \left(a_P + \frac{\bar{\alpha}z - \beta}{\bar{\beta}z + \alpha} 1_P\right) \otimes z'(\bar{\alpha} + \beta\bar{z})1_P.
\end{aligned}$$

The transformation of the conformal plane, corresponding to (3.5), may thus be written

$$z \to \frac{\bar{\alpha}z - \beta}{\bar{\beta}z + \alpha}. \tag{3.6}$$

It is well known that a complex transformation (3.6) corresponds to a rotation of the Riemann sphere.[7] Hence, (3.3) is an expression for the ordinary relation between spinors and the 2-sphere.

4. THE TANGENT VECTOR CORRESPONDING TO A NORMED SPINOR

Before proceeding, it should be mentioned that the methods of studying the unitary transformation (3.5) also apply to arbitrary, linear transformations of spinor space and to the corresponding full Lorentz group acting on the 2-sphere. For the sake of clarity, we will, at the present, only focus our attention on the most basic properties of spinors. We have therefore postponed this relativistic version of our spinor picture to a separate paper. Accordingly, instead of discussing the full connection between spinors and null flags, we restrict our study to the fundamental case of normed spinors[8] and normed tangent vectors to the 2-sphere. We first mention a general method to obtain this relation and then apply it to the specific case of $W_P \otimes \mathbf{C}_P$ and S^2.

Let us use the standard abbreviations SU_2 and SO_3 for the group of unimodular, unitary spinor transformations and the group of rotations of the 2-sphere, respectively. Note that a rotation r_{SO_3} also acts on the tangent bundle of the 2-sphere: each tangent vector n_Q will be translated along a circle of transitivity, the angle between n_Q and the circle remaining the same throughout the translation. In this way normed tangent vectors will be mapped into normed tangent vectors by r_{SO_3}. Normed spinors are trivially transformed into normed spinors under the action of unitary transformations r_{SU_2}. More subtly, however, there corresponds to any pair ζ, η of normed spinors precisely one transformation r_{SU_2} such that $r_{SU_2}\zeta = \eta$. Adopting the representation of unitary transformations by rotations $r_{SU_2} \to r_{SO_3}$, we may thus map normed spinors on normed tangent vectors in the following way: Select a single normed spinor ζ as a reference and relate ζ arbitrarily to some fixed normed tangent vector n_P (where P is a fixed point on the 2-sphere). Now, to any arbitrary normed spinor η there is precisely one r_{SU_2} such that $r_{SU_2}\zeta = \eta$. To the transformation r_{SU_2} there corresponds a rotation r_{SO_3} and, letting r_{SO_3} act on n_P, we obtain a new normed tangent vector, which thus has been related to η. Clearly, the relation between normed spinors and normed tangent vectors, so obtained, is invariant under unitary transformations and simultaneous rotations. Moreover, varying the tangent vector n_p, related to the fixed spinor ζ, every invariant representation of normed spinors by tangent vectors can be found.

In our particular spinor space model, normed spinors have the form $x_P^N \otimes e_P^{i\alpha}$, where x_P^N is a unit vector of W_P [cf. (3.1)]. Whenever x_P^N belongs to the interior of W_P we may alternatively write

$$x_P^N \otimes e_P^{i\alpha} = PQ^N \otimes e_P^{i\alpha},$$

thus making the projection $Q \in S^2$ of $x_P^N \otimes e_P^{i\alpha}$ evident. For reference spinor ζ in $W_P \otimes \mathbf{C}_P$ we choose $1_P \otimes 1_P$. We are free to relate any normed tangent vector to this spinor, but the obvious choice is of course 1_P. Next, we must characterize the circles of transitivity on S^2, corresponding to the unitary transformation $1_P \otimes 1_P \to PQ^N \otimes e_P^{i\alpha}$. Note that to any tangent vector $e_P^{i\gamma}$ at P and any other point $Q' \in S^2$ there is one and only one oriented circle $\Sigma(e_P^{i\gamma}, Q')$ on S^2 to which $e_P^{i\gamma}$ is a tangent and which contains Q'. As we shall see, the circles of transitivity are those parallel to $\Sigma(e_P^{i\alpha}, Q)$.

The unitary transformation $1_P \otimes 1_P \to PQ^N \otimes e_P^{i\alpha}$ may be written in full as the basis transformation

$$\begin{aligned}
&a_P \otimes 1_P \to P\dot{Q}^N \otimes e_P^{i(\psi - \alpha)}, \\
&1_P \otimes 1_P \to PQ^N \otimes e_P^{i\alpha}.
\end{aligned} \tag{4.1}$$

Here, for any point $Q' \in S^2$, $\dot{Q}' \in S^2$ denotes the point diametrically opposite Q (thus $a_p = p\dot{p} = P\dot{P}^N$). The angle ψ is the direction of PQ^N in $\mathbf{C}_P$:

$$PQ^N = \rho_1 a_p + \rho_2 e_P^{i\psi}$$

(where $\rho_1, \rho_2 \in \mathbf{R}_+$). The reader should check that, in rewriting (4.1) by means of (3.1) and (3.2), a matrix of the form

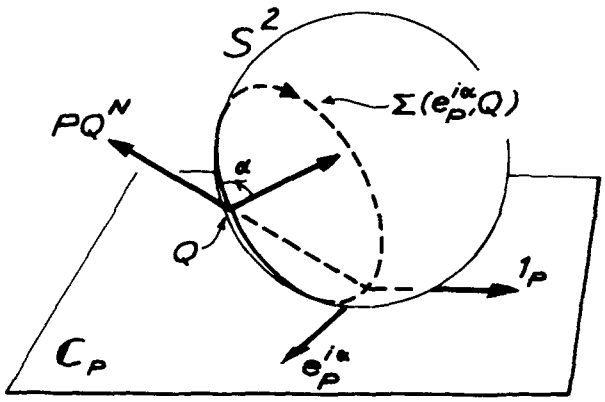

FIG. 3. Associated to the normed spinor $PQ^N \otimes e_P^{i\alpha}$, there is a normed tangent vector at Q, forming the angle α with $\Sigma(e_P^{i\alpha},Q)$. For clarity, in this and the following figures tangent vectors are deliberately not drawn to their full length.

(3.5) is obtained. In order to verify that (4.1) transforms the circle $\Sigma(e_P^{i\alpha},Q)$ into itself we regard spinors of the form

$$(\rho_1 PQ^N + \rho_2 e_P^{i\alpha}) \otimes e_P^{i\alpha}, \tag{4.2}$$

where $\rho_1 \in \mathbf{R}_+$ while $\rho_2 \in \mathbf{R}$ is arbitrarily real. Clearly, the projection of the spinors (4.2) on S^2 spans $\Sigma(e_P^{i\alpha},Q)$. Moreover, (4.2) may be rewritten

$$\rho_1(PQ^N \otimes e_P^{i\alpha}) + \rho_2(1_P \otimes 1_P).$$

Computing as in ordinary component-dependent spinor algebra it is straightforward to verify that this expression remains invariant in form under (4.1). The assertion is thus proved. It follows that the tangent vector 1_P moves from P to Q along $\Sigma(e_P^{i\alpha},Q)$ under the unitary transformation taking $1_P \otimes 1_P$ into $PQ^N \otimes e_P^{i\alpha}$. In other words, *the normed spinor $PQ^N \otimes e_P^{i\alpha}$ becomes associated to the normed tangent vector at Q, forming the angle α with $\Sigma(e_P^{i\alpha},Q)$* (cf. Fig. 3).

It is interesting to note that, changing the "phase" $e_P^{i\alpha} \to e_P^{i\gamma}$ in the spinor $PQ^N \otimes e_P^{i\alpha}$ and changing the circle $\Sigma(e_P^{i\alpha},Q)$ accordingly, the tangent vector at Q, obtained by translating 1_P along $\Sigma(e_P^{i\alpha},Q)$, is rotated through *twice* the angle $\gamma - \alpha$. This is, so to speak, the material reason for the well-known double degeneracy in the relationship between spinors and tangent vectors. Translating 1_P along the opposite routes of $\Sigma(\pm e_P^{i\alpha},Q)$, we find that one and the same tangent vector precisely corresponds to the two spinors $PQ^N \otimes \pm e_P^{i\alpha}$, differing only in sign.

We have not yet discussed spinors of the form $e_P^{i\psi} \otimes e_P^{i\alpha}$. In this limit we may replace the circle $\Sigma(e_P^{i\alpha},Q)$ by a circle in $\mathbf{C}_P$ for which $e_P^{i\alpha}$ is a tangent and on which $e_P^{i\psi}$ is a point. It is not difficult to see that translation of 1_P along this circle yields a tangent vector $e_P^{i2(\psi - \alpha)}$. The consistency between this result and (C) should be noted in particular.

5. TRANSFORMATION BETWEEN HALF-SPACES

Two planes $\mathbf{C}_P$, $\mathbf{C}'_Q$ could differ from each other by being tangent planes to two different points of S^2 or by being the same tangent plane at $P = Q \in S^2$ but possessing different unit directions $1_P \in \mathbf{C}_P$, $1'_P \in \mathbf{C}'_P$. In this paragraph it will be established that, to any normed spinor $PQ^N \otimes e_P^{i\alpha}$ of $W_P \otimes \mathbf{C}_P$, there is associated a space $W_Q \otimes \mathbf{C}'_Q$ and an isomorphism transforming $PQ^N \otimes e_P^{i\alpha}$ into the "canonical" spinor $1'_Q \otimes 1'_Q$ of $W_Q \otimes \mathbf{C}'_Q$.

The isomorphism

$$W_P \otimes \mathbf{C}_P \sim W_Q \otimes \mathbf{C}'_Q \tag{5.1}$$

we are about to describe, resembles a "passive" transformation of ordinary vector calculus. One way of bringing it about is to write

$$\begin{aligned} &PQ^N \otimes e_P^{i\alpha} \sim 1'_Q \otimes 1'_Q, \\ &P\dot{Q}^N \otimes e_P^{i(\psi - \alpha)} \sim Q\dot{Q} \otimes 1'_Q \end{aligned} \tag{5.2}$$

($1'_Q$ is the tangent vector corresponding to $PQ^N \otimes e_P^{i\alpha}$; ψ is the direction of PQ^N in $\mathbf{C}_P$), and extend this pairing by linearity to the entire of $W_P \otimes \mathbf{C}_P$ and $W_Q \otimes \mathbf{C}'_Q$. The tangent vectors associated to normed spinors of the two spaces

$$\alpha(P\dot{P} \otimes 1_P) + \beta(1_P \otimes 1_P)$$

and

$$\alpha(Q\dot{Q} \otimes 1'_Q) + \beta(1'_Q \otimes 1'_Q)$$

($\alpha, \beta \in \mathbf{C}$) can, because of the symmetry of S^2, differ only by the same rotation which transforms 1_P into $1'_Q$. Since this rotation also corresponds to the unitary transformation (4.2), we find that *spinors paired according to* (5.2) *correspond to the same tangent vector.* It should be clear that the form of the transformation (5.1) is essentially (in fact up to a sign) determined from this invariance. We may therefore *start* from the correspondence between spinors and tangent vectors (and the geometrical description given to it) in order to obtain a more geometrical description of the transformation than its purely component-dependent definition (5.2).

It will be convenient to consider first the case when $P = Q$ in (5.1). Let us say that the unit directions of $\mathbf{C}_P$ and $\mathbf{C}'_P$ differ according to $1'_P = e_P^{i\psi}$. We wish to pair normed spinors $x_P^N \otimes e_P^{i\alpha}$ and $y_P^N \otimes (e')_P^{i\beta}$ [where $(e')_P^{i\beta} = e_P^{i(\beta + \psi)}$] in such a way that they correspond to the same tangent vector. Thus, for normed spinors, the pairing must be of the form

$$PQ^N \otimes e_P^{i\alpha} \sim PQ^N \otimes (e')_P^{i\beta}. \tag{5.3}$$

Suppose that translation of 1_P along $\Sigma(e_P^{i\alpha},Q)$ and translation of $1'_P$ along $\Sigma((e')_P^{i\beta},Q) = \Sigma(e_P^{i(\beta + \psi)},Q)$ yield tangent vectors n_Q, n'_Q at Q, respectively. Defining δ as the angle at P between the two intersecting circles

$$\Sigma((e')_P^{i\beta},Q) = \Sigma(e_P^{i(\alpha + \delta)},Q)$$

the angle between $\Sigma((e')_P^{i\beta},Q)$ and n_Q becomes $\alpha - \delta$ (see Fig. 4). With this notation, in order that $n_Q = n'_Q$ we must have

$$\alpha - \delta = \beta = \alpha + \delta - \psi,$$

i.e., $2\delta = \psi$. Accordingly, we are left with two possible choices for the transformation (5.3):

$$PQ^N \otimes e_P^{i(\beta + \psi/2)} \sim PQ^N \otimes (e')_P^{i\beta} \tag{5.4}$$

or

$$PQ^N \otimes -e_P^{i(\beta + \psi/2)} \sim PQ^N \otimes (e')_P^{i\beta}. \tag{5.5}$$

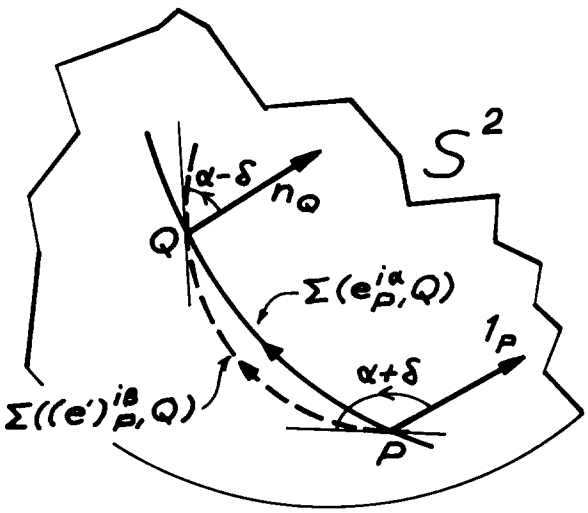

FIG. 4. The tangent vectors 1_P and n_Q both form the same angle α with $\Sigma(e_P^{i\alpha},Q)$. When changing this circle into $\Sigma((e')_P^{i\beta},Q)$, the angle α changes by the same amount δ but in opposite directions at P and Q.

In the limit $Q\to P$ these expressions become:

$$e_P^{i\psi}\otimes e_P^{i\psi/2}=1_P\otimes e_P^{-i\psi/2}\sim 1'_P\otimes 1'_P, \tag{5.4a}$$

$$e_P^{i\psi}\otimes -e_P^{i\psi/2}=1_P\otimes -e_P^{-i\psi/2}\sim 1'_P\otimes 1'_P, \tag{5.5a}$$

respectively. We will say that (5.3) and (5.4) are the unique transformations mapping each of the spinors $1_P\otimes \pm e_P^{-i\psi/2}$ into canonical form.

Allowing $P\neq Q$ in (5.1), we may pick out any normed spinor $PQ^N\otimes e_P^{i\alpha}$ to be transformed into canonical form:

$$PQ^N\otimes e_P^{i\alpha}\sim 1_Q\otimes 1_Q \tag{5.6}$$

(since $P\neq Q$ no inconsistency will arise if we drop the prime in C'_Q). The pairing (5.6) can be extended in a unique way to a transformation (5.1), just as (5.4) and (5.5) are the unique of extensions of (5.4a) and (5.5a). From the invariant correspondence between spinors and tangent vectors, it is clear that the transformation must be of the form

$$PR^N\otimes e_P^{i(\gamma+\varphi)}\sim QR^N\otimes e_Q^{i\gamma},$$

where (5.6) is the particular case $R\to Q$. In general, the angle φ will depend on $R\in S^2$ and we shall now determine this dependence. We introduce the unique, oriented circle on S^2 passing through the points P,Q,R in turn:

$$\Sigma(P,Q,R)=\Sigma(e_P^{i(\alpha+\delta)},Q) \tag{5.7}$$

thus defining δ as the angle between $\Sigma(P,Q,R)$ and $\Sigma(e_P^{i\alpha},Q)$ at P (cf. Fig. 5). For the point R, we first seek the transformation mapping $PQ^N\otimes e_P^{i(\alpha+\delta)}$ into canonical form (denoting by $1'_Q$ its associated tangent vector)

$$PQ^N\otimes e_P^{i(\alpha+\delta)}\sim 1'_Q\otimes 1'_Q. \tag{5.8}$$

Since 1_P and $1'_Q$ form the same angle with $\Sigma(P,Q,R)$, they end up in the same tangent vector at R when translated along this circle. Hence, we may pair spinors in any of the two ways

$$PR^N\otimes \pm e_P^{i(\alpha+\delta)}\sim QR^N\otimes (e')_Q^{i(\alpha+\delta)}.$$

Only for one choice of sign the pairing agrees with (5.8), however, as is seen letting $R\to Q$. In this limit $QR^N\to (e')_Q^{i(\alpha+\delta)}$, hence

$$PR^N\otimes e_P^{i(\alpha+\delta)}\sim QR^N\otimes (e')_Q^{i(\alpha+\delta)}. \tag{5.9}$$

From (5.6) and (5.8) we have $1_Q\otimes e_Q^{i\delta}\sim 1'_Q\otimes 1'_Q$, implying [cf. (5.4)]

$$QR^N\otimes (e')_Q^{i(\alpha+\delta)}\sim QR^N\otimes e_Q^{i\alpha},$$

which may be inserted in (5.9),

$$PR^N\otimes e_P^{i(\alpha+\delta)}\sim QR^N\otimes e_Q^{i\alpha}. \tag{5.10}$$

As has already been discussed in the last paragraph, changing the angle α to any other value γ, the tangent vector associated to the spinors in (5.10) becomes rotated through twice the angle $\gamma-\alpha$. The relation (5.10) will still hold and we may write

$$PR^N\otimes e_P^{i(\gamma+\delta)}\sim QR^N\otimes e_Q^{i\gamma}. \tag{5.11}$$

Formula (5.11) *is the general expression for the transformation of normed spinors induced by* (5.6). The angle δ as a function of $R\in S^2$ is defined by (5.7)

So far we have not dealt with nonnormed spinors. We extend (5.11) to this class linearly, writing

$$PR^N\otimes \rho e_P^{i(\gamma+\delta)}\sim QR^N\otimes \rho e_Q^{i\gamma}, \tag{5.11a}$$

where $\rho\in\mathbf{R}_+$. As was discussed at the beginning of this paragraph, pairing spinors which correspond to the same tangent vector gives the linear transformation (5.2). Consequently, (5.11) [or (5.11a)] is a *complex-linear* transformation and must in fact agree with (5.2). Let us, as a check, apply (5.11) to the point $R=Q$. In this case the plane of the circle $\Sigma(P,Q,R)=\Sigma(P,Q,Q)$ is orthogonal to C_P and we may write

$$\Sigma(P,Q,Q)=\Sigma(e_P^{i\psi},Q),$$

where ψ is the angle of PQ^N in the plane C_P [cf. (4.2)]. Consequently $\delta=\psi-\alpha$, and (5.11) takes the form

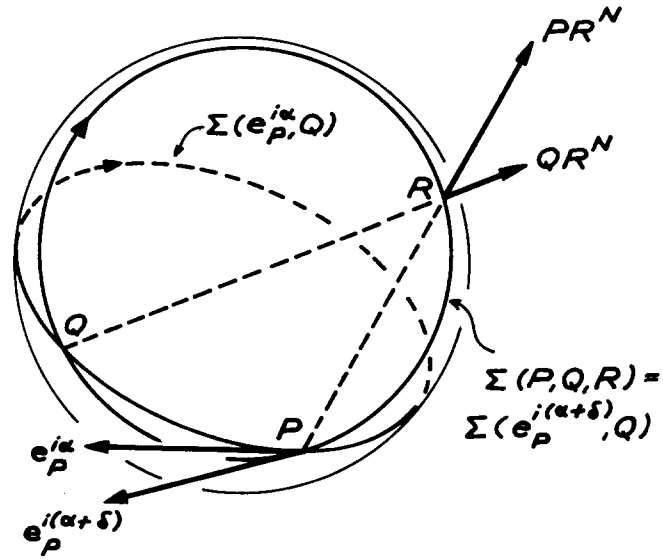

FIG. 5. Interpretation of the phase shift δ in the transformation $PR^N\otimes e_P^{i(\gamma+\delta)}\to QR^N\otimes e_Q^{i\gamma}$. The unit directions 1_P, 1_Q (not shown) are supposed to form the same angle with $\Sigma(e_P^{i\alpha},Q)$. To obtain a precise definition of δ, we rewrite $\Sigma(P,Q,R)$ in terms of its tangent vector at P: $\Sigma(P,Q,R)=\Sigma(e_P^{i(\alpha+\delta)},Q)$.

$$P\dot{Q}^N \otimes e_P^{i(\psi-\alpha)} \sim Q\dot{Q} \otimes 1_Q. \tag{5.12}$$

Evidently, the two lines of (5.2) is obtained by letting $R \to Q$ [cf. (5.6)] and $R = \dot{Q}$ in (5.11).

6. THE GEOMETRICAL INTERPRETATION OF SPINOR ADDITION

It will be demonstrated that for any pair of spinors

$$x_P \otimes z_P, y_P \otimes u_P \in W_P \otimes \mathbf{C}_P$$

$(x_P \neq y_P, z_P \neq u_P)$ there is a space $W_Q \otimes \mathbf{C}_Q$ in which these correspond to a pair of spinors

$$x'_Q \otimes z'_Q,\ y'_Q \otimes z'_Q \in W_Q \otimes \mathbf{C}_Q,$$

both possessing the same phase z'_Q. Thus, we may geometrically obtain the sum of two spinors, passing to this particular space and there form their sum by means of the parallelogram rule applied to the vectors x'_Q, y'_Q [cf. rule (A)].

As a preliminary step we study the transformation (5.11) in the limit $PR^N \to e_P^{i\beta}$ where $e_P^{i\beta}$ is any direction in $\mathbf{C}_P$. In this limit $\Sigma(P,Q,R) \to \Sigma(-e_P^{i\beta}, Q)$ so the angle δ of (5.7) becomes $\pi + \beta - \alpha$. Accordingly,

$$e_P^{i\beta} \otimes -e_P^{i(\beta-\alpha)} \sim QP^N \otimes 1_Q,$$

or deleting the angle β,

$$1_P \otimes 1_P \sim QP^N \otimes -e_Q^{i\alpha}, \tag{6.1}$$

which is the required formula. Now consider two circles on S^2:

$$\Sigma(P,Q,R): 1_Q \leftrightarrow 1_R,$$
$$\Sigma(P,Q,R_0): 1_P \leftrightarrow 1_Q \leftrightarrow 1_{R_0},$$

translating the tangent vectors 1_P, 1_Q, 1_R, 1_{R_0}, into each other in the manner indicated. Associated to these circles are transformations (5.1) such that

$$\begin{aligned} PR_0^N \otimes e_P^{i\delta} &\sim QR^N \otimes 1_Q \sim 1_R \otimes 1_R, \\ PR_0^N \otimes 1_P &\sim QR_0^N \otimes 1_Q \sim 1_{R_0} \otimes 1_{R_0}. \end{aligned} \tag{6.2}$$

Here, δ is the angle between $\Sigma(P,Q,R)$ and $\Sigma(P,Q,R_0)$ at P. In order to define the sign of δ, we reexpress the two circles in terms of their tangent vectors at P:

$$\begin{aligned} \Sigma(P,Q,R) &= \Sigma(e_P^{i(\alpha+\delta)}, Q), \\ \Sigma(P,Q,R_0) &= \Sigma(e_P^{i\alpha}, Q). \end{aligned} \tag{6.3}$$

Comparing (5.6) and (6.1), we find that (6.2) implies

$$\begin{aligned} &1_P \otimes 1_P \sim RP^N \otimes -e_R^{i\delta} \sim R_0P^N \otimes -1_{R_0}, \\ &1_Q \otimes 1_Q \sim RQ^N \otimes -1_R \sim R_0Q^N \otimes -1_{R_0}. \end{aligned} \tag{6.4}$$

Let us regard a sum of two spinors in $W_R \otimes \mathbf{C}_R$:

$$RP^N \otimes \rho_1 e_R^{i(\gamma+\delta)} + RQ^N \otimes \rho_2 e_R^{i\gamma}, \tag{6.5}$$

where $\rho_1, \rho_2 \in \mathbf{R}_+$. According to (6.4) the expression (6.5) is equivalent to

$$(\rho_1 R_0 P^N + \rho_2 R_0 Q^N) \otimes e_{R_0}^{i\gamma}, \tag{6.6}$$

defined in $W_{R_0} \otimes \mathbf{C}_{R_0}$. Formula (6.6) is the general expression for the sum of two spinors with different directions in W_R and different phases. As indicated by (6.3), *such a sum may be evaluated in a component-independent way if we pass to any space $W_{R_0} \otimes \mathbf{C}_{R_0}$ for which the circle $\Sigma(P,Q,R_0)$ differs from $\Sigma(P,Q,R)$ by an angle equal to the phase difference δ.* In $W_{R_0} \otimes \mathbf{C}_{R_0}$ we may apply the parallelogram rule to find the sum. The situation is thus as illustrated in Fig. 6.

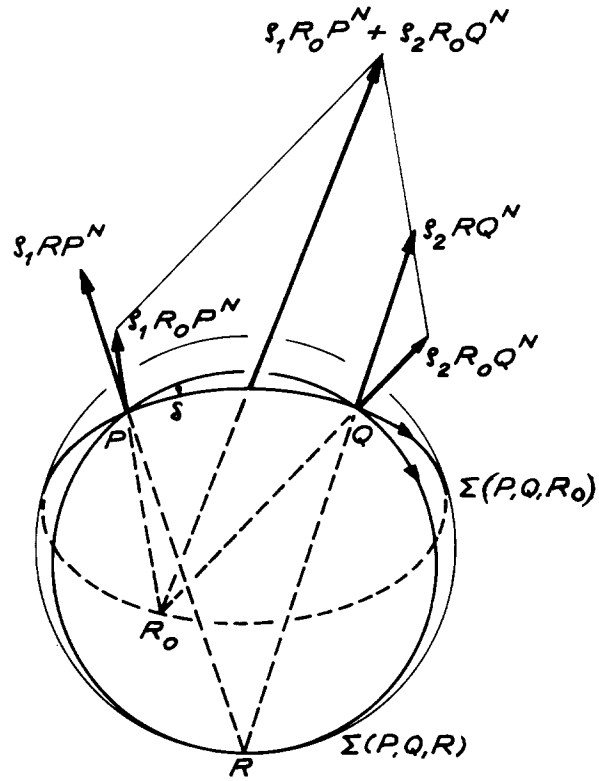

FIG. 6. Geometrical interpretation of the sum $RP^N \otimes \rho_1 e_R^{i(\gamma+\delta)} + RQ^N \otimes \rho_2 e_R^{i\gamma}$. First we pass to the equivalent expression $R_0P^N \otimes e_{R_0}^{i\gamma} + R_0Q^N \otimes \rho_2 e_{R_0}^{i\gamma}$. To this the parallelogram rule is applied. The sum is thus found to be $(\rho_1 R_0 P^N + \rho_2 R_0 Q^N) \otimes e_{R_0}^{i\gamma}$.

The reader is well-advised to study this procedure of spinor addition in detail. Note, in particular, that the point R_0 of (6.6) may be replaced by any other point $R'_0 \in \Sigma(P,Q,R_0)$ giving the same orientation to the circle, i.e.,

$$\Sigma(P,Q,R_0) = \Sigma(P,Q,R'_0).$$

The projection of (6.6) on S^2 is invariant under this substitution. However, if we, instead of R_0 choose a point $R''_0 \in \Sigma(P,Q,R_0)$ giving the opposite orientation to the circle, i.e.,

$$\Sigma(P,Q,R_0) = \Sigma(P,R''_0,Q)$$

the space $W_{R''_0} \otimes \mathbf{C}_{R''_0}$ will correspond to the *difference* between the two spinors. The projection of this difference on S^2 will also be located on $\Sigma(P,Q,R_0)$ but on the side of P and Q, which is opposite to the projection of the sum.

ACKNOWLEDGMENT

The author would like to thank Professor B. Laurent for many useful comments.

[1] N.R. Campbell, *Physics, the Elements* (Cambridge, England, 1920). Reprinted (Dover, New York, 1957) and entitled *The Foundations of Science*.
[2] E. Cartan, *The Theory of Spinors* (Hermann, Paris, 1966).

[3]O. Veblen, Proc. Natl. Acad. Sci. **19**, 462 (1933).

[4]R. Penrose in *Batelle Rencontres*, 1967 Lectures in Mathematics and Physics, edited by C.M. DeWitt and J.A. Wheeler (Benjamin, New York, 1968).

[5]Actually, it is possible *in principle* to use the null flag picture to interpret algebraic operations on spinors. However, the corresponding operations on null flags will be quite complicated to handle (which is not very surprising, considering the two-to-one correspondence between spinors and null flags). An investigation of this kind has been carried out by R. Penrose (unpublished).

[6]One may check that this projection, suitably restricted, has the formal properties of the so-called Hopf map, which really is the core of the relation between spinors and world vectors.

[7]I.M. Gelfand, R.A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, Oxford, 1963).

[8]By normed spinors we of course mean spinors $(z_1, z_2) \in \mathbf{C}^2$, subjected to the restriction $|z_1|^2 + |z_2|^2 = 1$.

Disequilibrium theory applied to two-spin Glauber model [a)]

Hwe Ik Zhang and Jae Il Lee
Department of Physics, Seoul National University, Seoul 151, Korea

(Received 6 June 1979; accepted for publication 22 August 1979)

The macroscopic disequilibrium theory for Markovian relaxation processes is applied to the two-spin Glauber model. The theory is tested rigorously by evaluation of all quantities explicitly. The probability distribution function is shown to be reproduced exactly by the minimal information procedure using the knowledge of certain macroscopic variables. The probability distribution functions obtained using a reduced number of macroscopic observables are examined and the deviations from the exact distribution function are discussed. It is concluded that the disequilibrium situation might be described satisfactorily if we choose a suitable set of macroscopic variables, but the result might crucially depend on the choice of these variables. The entropy deficiency, entropy production, and the first and the second time derivatives of the entropy production are evaluated explicitly for this model. The signs of such quantities are found to be in accordance with the predictions of more qualitative theories.

I. INTRODUCTION

A theory dealing with macroscopic disequilibrium for Markovian relaxation processes has recently been developed by Levine and co-workers,[1-3] and others.[4,5] This theory relates the probability distribution functions to a few macroscopic observables via the minimal information procedure, introducing intensive thermodynamic quantities ("forces") as the Lagrange multipliers. The time derivatives of these macroscopic observables ("fluxes") and the corresponding "forces" provide a legitimate connection to the irreversible thermodynamics developed by Onsager,[6] and Prigogine and co-workers.[7,8] Some elementary applications of this theory has been made to the case of vibrational relaxation of diatomic molecules in a buffer gas. [1]

It is desirable, however, to find a situation in which the theory can be tested rigorously by evaluating all quantities involved explicitly. One of the simplest model system for such a situation is provided by the Glauber model.[9] The probability distribution functions of a finite Ising chain can be obtained precisely by a procedure of correlation expansion in the Glauber model.

The main purposes of the present work are as follows. We, firstly, apply the disequilibrium theory to the system of which the exact solution to the master equation, or the time-dependent probability function, is known. We intend to verify that the nonequilibrium situation is described exactly by the knowledge of a set of macroscopic observables. We also study the deviation in system description due to the reduced number of macroscopic observables employed in the minimal information procedure. It is possible to evaluate the entropy deficiency, entropy production, and the first and second time derivatives of the entropy production for our model. This provides a very rare opportunity to verify the prediction of more qualitative theories such as irreversible thermodynamics concerning the signs of such quantities.

The arrangements of this paper is as follows. We review the disequilibrium theory in Sec. II. We show clearly the relation between the entropy production and the net time rate of entropy produced in reservoir and system in a relaxation process. The exact solutions to the average spin and spin correlation function in two-spin Glauber model are obtained in Sec. III. We apply explicitly the disequilibrium theory to the Glauber model in Sec. IV. We obtain the expressions for the probability distribution functions and "forces" with sufficient and reduced number of macroscopic observables. We compare the results with the exact values. The entropy deficiency, entropy production, and the first and second time derivatives are calculated explicitly and discussed in relation to the inequality of Prigogine, in Sec. V. Section VI presents our conclusions.

II. DISEQUILIBRIUM THEORY

We consider a system in contact to a heat reservoir, in which the Markovian relaxation process is described by a master equation,

$$\frac{d}{dt}P(n,t) = -\sum_{n'} P(n,t)W_{nn'} + \sum_{n'} P(n',t)W_{n'n}, \quad (1)$$

where $P(n,t)$ is the probability for the system to be found in a state n at time t, and $W_{nn'}$ is the transition rate from state n to state n'. The probability distribution function $P(n,t)$ satisfy the obvious normalization condition

$$\sum_{n} P(n,t) = 1. \quad (2)$$

The disequilibrium entropy of the system at time t can be defined in the form of missing information

$$S(t) = -\sum_{n} P(n,t)\ln P(n,t). \quad (3)$$

We now introduce a set of macroscopic observables $\{A_r, r = 1,2...,M\}$ which is regarded to be sufficient to characterize the disequilibrium situation macroscopically. The time dependence of these variables is expressed by the relation

$$\langle A_r(t)\rangle = \sum_{n} P(n,t)A_r(n), \quad (4)$$

where $A_r(n)$ is the value of A_r at a particular state n. Once we know the probability distribution function assuming, for ex-

[a)]Supported by the Research Institute for Basic Sciences, Seoul National University.

 0022-2488/79/122438-07$01.00

ample, Eq. (1) is solved, the disequilibrium entropy and macroscopic observables are determined by Eqs. (3) and (4).

It might be desirable, however, to estimate the probability distribution functions using the phenomenological knowledge of $A_r(t)$. To get the best estimation, we require a minimum information procedure which maximizes $S(t)$ under the constraints of Eqs. (2) and (4). If we introduce the Lagrange multipliers $\lambda_0(t)$ and $\lambda_r(t)$ $[r = 1,2,..., M]$ corresponding to the constraints Eqs. (2) and (4), we find,

$$P(n,t) = \exp\Big(-\lambda_0(t) - \sum_{r=1}^{M} \lambda_r(t) A_r(n) \Big). \tag{5}$$

The Lagrange multipliers $\lambda_0(t)$ and $\lambda_r(t)$ are, in principle, determined by the conditions

$$\sum_n \exp\Big(-\lambda_0(t) - \sum_{r=1}^{M} \lambda_r(t) A_r(n) \Big) = 1, \tag{6}$$

$$\sum_n A_{r'}(n) \exp\Big(-\lambda_0(t) - \sum_{r=1}^{M} \lambda_r(t) A_r(n) \Big) = \langle A_{r'}(t) \rangle \tag{7}$$

$(r' = 1,2,...,M)$.

Further macroscopic observables $\langle A_{r'}(t) \rangle$ $[r' = M+1,...]$ can be predicted using the probability distribution functions in Eq. (5) once the Lagrange multipliers are determined. The predicted observables, of course, may not coincide with the measured values if such measurements are carried out. But further information concerning such observables might be used to obtain better estimates for the probability distribution functions. This procedure of improving the knowledge of $P(n,t)$, however, may not continue indefinitely. After using the information contained in a certain number of observables, the estimate for $P(n,t)$ may no longer improve if all the other information is redundant.[1] For this case the Lagrange multipliers corresponding to the additional observables are regarded as vanishing. We expect that many Lagrange multipliers $\lambda_r(t)$ becomes vanishingly small as the system approaches equilibrium, since at equilibrium only a few observables are required to determine the probability distribution functions.

Using the expression $P(n,t)$ given in Eq. (5), the disequilibrium entropy can be expressed

$$S(t) = \lambda_0(t) + \sum_{r=1}^{M} \lambda_r(t) \langle A_r(t) \rangle. \tag{8}$$

The Lagrangian multipliers $\lambda_r(t)$ are the partial derivatives of $S(t)$ with regard to $\langle A_r(t) \rangle$:

$$\lambda_r(t) = \frac{\partial S(t)}{\partial \langle A_r(t) \rangle}. \tag{9}$$

The time derivative of disequilibrium entropy of the system

$$\frac{d}{dt} S(t) = -\sum_n [\ln P(n,t) + 1] \frac{d}{dt} P(n,t) \tag{10}$$

can be written, using Eqs. (2) and (5),

$$\frac{d}{dt} S(t) = \sum_{r=1}^{M} \lambda_r(t) \frac{d}{dt} \langle A_r(t) \rangle. \tag{11}$$

The equilibrium probability distribution $P^0(n)$ can also be written in the form of Eq. (5):

$$P^0(n) = \exp\Big[-\lambda_0^0 - \sum_{r=1}^{M} \lambda_r^0 A_r(n) \Big] \tag{12}$$

and the corresponding entropy S^0 is

$$S^0 = \lambda_0^0 + \sum_{r=1}^{M} \lambda_r^0 \langle A_r \rangle \tag{13}$$

understanding that many of the λ_r^0 vanishes. The observables corresponding to nonvanishing λ_r^0 are conserved quantities, such as energy, otherwise the entropy $S(t)$ in Eq. (8) can continue to increase by changing $\langle A_r(t) \rangle$. The time derivative of the entropy can now be written

$$\frac{d}{dt} S(t) = \sum_{r=1}^{M} \mu_r(t) \frac{d}{dt} \langle A_r(t) \rangle + \sum_{r=1}^{M} \lambda_r^0 \frac{d}{dt} \langle A_r(t) \rangle, \tag{14}$$

where

$$\mu_r(t) = \lambda_r(t) - \lambda_r^0. \tag{15}$$

The second term in Eq. (14) is precisely the negative time derivative of the entropy in the reservior:

$$\sum_{r=1}^{M} \lambda_r^0 \frac{d}{dt} \langle A_r(t) \rangle = -\sum_{r=1}^{M} \lambda_r^0 \frac{d}{dt} \langle A_r(t) \rangle_R = -\frac{d}{dt} S_R \tag{16}$$

since the observables A_r are conserved quantity

$$\frac{d}{dt} [\langle A_r(t) \rangle + \langle A_r(t) \rangle_R] = 0$$

unless $\lambda_r^0 = 0$, where the quantities with subscript R belong to the reservoir. The first term in Eq. (14) is therefore the net time rate of entropy due to the relaxation process in the system plus environment, which is usually call the "entropy production." It is very easily verified that this entropy production is the negative time derivative of entropy deficiency which is defined as the averaged deviance of $P(n,t)$ from stationary distribution $P^0(n)$,

$$D(t) = \sum_n P(n,t) \ln[P(n,t)/P^0(n)]. \tag{17}$$

To obtain the best probability distribution function, we may start from the entropy deficiency and follow minimizing procedure. The entropy production $\mathscr{P}(t)$ due to the relaxation process is now written,

$$\mathscr{P}(t) = -\frac{d}{dt} D(t) = \sum_{r=1}^{M} \mu_r(t) X_r(t) \tag{18}$$

in terms of the "forces" $\mu_r(t)$ given in Eq. (15) and "fluxes" $X_r(t)$ defined by

$$X_r(t) = \frac{d}{dt} \langle A_r(t) \rangle. \tag{19}$$

It has been shown[2,4] that both the entropy deficiency and entropy production are non-negative quantities and this is equivalent to the second law of thermodynamics.

III. TWO-SPIN GLAUBER MODEL

The kinetic Ising model, which was introduced by Glauber,[9] relates the strongly interacting Ising model to a time-dependent disequilibrium situation via a master equation in the form of Eq. (1). In this model the scalar spins,

which are located on lattice sites, can take the values ± 1 corresponding to the up and down positions of the spins as in an Ising model. For the simplicity of our model, we take an Ising chain with only two spins surrounded by a heat reservoir.

The Hamiltonian of the two-spin Ising chain is

$$H = -J\sigma_1\sigma_2, \tag{20}$$

where σ_1,σ_2 are the spin values and J is the interaction strength between these two spins. We write the probability distribution function in the form $P(\sigma_1,\sigma_2,t)$ for each set of (σ_1,σ_2) at time t. If we let $\omega_j(\sigma_1,\sigma_2)$ be the probability per unit time that the jth spin flips from σ_j to $-\sigma_j$ when the initial spin complexion is (σ_1,σ_2), the master equation can be written

$$\begin{aligned}\frac{d}{dt}P(\sigma_1,\sigma_2,t) = &-\omega_1(\sigma_1,\sigma_2)P(\sigma_1,\sigma_2,t)\\ &-\omega_2(\sigma_1,\sigma_2)P(\sigma_1,\sigma_2,t)\\ &+\omega_1(-\sigma_1,\sigma_2)P(-\sigma_1,\sigma_2,t)\\ &+\omega_2(\sigma_1,-\sigma_2)P(\sigma_1,-\sigma_2,t).\end{aligned} \tag{21}$$

To determine the possible form of $\omega_j(\sigma_1,\sigma_2)$, we allow the tendency for the spins to align parallel each other and choose the form, as Glauber did,

$$\omega_j(\sigma_1,\sigma_2) = \tfrac{1}{2}\alpha(1-\eta\sigma_1\sigma_2) \tag{22}$$

which would have two possible values, $\frac{1}{2}\alpha(1-\eta)$ and $\frac{1}{2}(1+\eta)$.

The parameter α is related to the interaction strength between the system and the reservoir. Since this parameter describes the time scale on which all transitions take place, we let it be unity for conveience. The other parameter η, describes the tendency toward alignment and describes the equilibrium state of system. We therefore choose η in such a way that the system would reduce to the correct equilibrium distribution. Following the procedure provided by Glauber, we obtain

$$\eta = \tanh(J/kT). \tag{23}$$

The probability distribution functions are given by[9]

$$P(\sigma_1,\sigma_2,t) = \tfrac{1}{4}\Big(1 + \sum_j \sigma_j q_j(t) + \sum_{j\neq k}\sigma_j\sigma_k r_{j,k}(t)\Big), \tag{24}$$

where $q_j(t)$ and $r_{j,k}(t)$ are the expectation values of σ_j and $\sigma_j\sigma_k$ defined by

$$q_j(t) = \langle\sigma_j(t)\rangle = \sum_{\{\sigma\}}\sigma_j P(\sigma_1,\sigma_2,t), \tag{25}$$

$$r_{j,k}(t) = \langle\sigma_j(t)\sigma_k(t)\rangle = \sum_{\{\sigma\}}\sigma_j\sigma_k P(\sigma_1,\sigma_2,t), \tag{26}$$

where the summation is taken over all the configuration of spins. The spin expectation functions $q_j(t)$ satisfy the flowing coupled equations

$$\begin{aligned}\frac{d}{dt}q_1(t) &= -q_1(t) + \eta q_2(t),\\ \frac{d}{dt}q_2(t) &= -q_2(t) + \eta q_1(t),\end{aligned} \tag{27}$$

while the spin correlation functions $r_{ij}(t)$ satisfy

$$\begin{aligned}\frac{d}{dt}r_{12}(t) &= -2r_{12}(t) + 2\eta,\\ r_{12}(t) &= r_{12}(t).\end{aligned} \tag{28}$$

The solution of Eqs. (27) and (28) are easily found:

$$\begin{aligned}q_1(t) = &\tfrac{1}{2}\{q_1(0)+q_2(0)\}e^{-(1-\eta)t}\\ &+\tfrac{1}{2}\{q_1(0)-q_2(0)\}e^{-(1+\eta)t},\end{aligned} \tag{29}$$

$$\begin{aligned}q_2(t) = &\tfrac{1}{2}\{q_1(0)+q_2(0)\}e^{-(1-\eta)t}\\ &-\tfrac{1}{2}\{q_1(0)-q_2(0)\}e^{-(1+\eta)t},\end{aligned} \tag{30}$$

$$r(t) = \{r(0)-\eta\}e^{-2t} + \eta, \tag{31}$$

where $r(t)\equiv r_{12}(t) = r_{21}(t)$. Equation (24) with Eqs. (29)–(31) completely determine the probability distribution functions $P(\sigma_1,\sigma_2,t)$, once the initial conditions, $q_1(0)$, $q_2(0)$, and $r(0)$, are prescribed.

IV. APPLICATION OF DISEQUILIBRIUM THEORY TO GLAUBER MODEL

The two-spin Glauber model described in Sec. III provides an excellent situation in which the disequilibrium theory can be tested rigorously, since we have already exact solutions for the probability distribution functions to compare with the same distribution functions determined by the disequilibrium theory. We also have the spin expectation functions $q_j(t)$ and spin correlation function $r_{12}(t)$ evaluated exactly, which can be chosen as the macroscopic observables. We consider two different initial spin configurations: initially aligned and initially disaligned.

A. Initially aligned configuration

We assume that the initial values of the two spins are both $+1$, which means

$$q_1(0) = q_2(0) = 1$$

and

$$r(0) = 1.$$

We then have, from Eqs (29)–(31),

$$q_1(t) = q_2(t) = e^{-(1-\eta)t}, \quad r_{12}(t) = \eta + (1-\eta)e^{-2t}. \tag{32}$$

For a macroscopic variable it is more natural to choose the average spin function

$$\bar{q}(t) = \tfrac{1}{2}[q_1(t) + q_2(t)] \tag{33}$$

rather than $q_1(t)$ and $q_2(t)$ individually.

We have freedom to choose the set of macroscopic observables out of the independent expectation value functions $\bar{q}(t)$, $r(t)$, and $q_1(t)$. We therefore try the following four possibilities:

(I) $\bar{q}(t), r_{12}(t), q_1(t)$,

(II) $\bar{q}(t), r_{12}(t)$,

(III) $\bar{q}(t)$,

(IV) $r_{12}(t)$.

It turns out immediately that, in the observable set I, $q_1(t)$ is redundant, and therefore set I reduces to the observable set II. For the sake of notational simplification, we introduce

two functions

$$q \equiv q(t) = e^{-(1-\eta)t},$$
$$r \equiv r(t) = \eta + (1-\eta)e^{-2t} \tag{34}$$

which are indentical to $\bar{q}(t)$ and $r_{12}(t)$, respectively,

For the observable set II, we let

$$A_1 = \tfrac{1}{2}(\sigma_1 + \sigma_2), \quad \langle A_1(t)\rangle = q(t),$$
$$A_2 = \sigma_1\sigma_2, \quad \langle A_2(t)\rangle = r(t), \tag{35}$$

with corresponding Lagrange multipliers $\lambda_1(t)$ and $\lambda_2(t)$, which should be determined by the conditions given in Eq. (7). Also we introduce the convention

$$P(1,t) = P(1,1,t), \quad P(2,t) = P(1,-1,t),$$
$$P(3,t) = P(-1,1,t), \quad P(4,t) = P(-1,-1,t). \tag{36}$$

Then, the distribution obtained by the minimum information procedure, Eq. (5), is written

$$P(1,t) = \exp[-\lambda_0 - \lambda_1 - \lambda_2],$$
$$P(2,t) = \exp[-\lambda_0 + \lambda_2],$$
$$P(3,t) = \exp[-\lambda_0 + \lambda_2], \tag{37}$$
$$P(4,t) = \exp[-\lambda_0 + \lambda_1 - \lambda_2].$$

The equations determining λ_0, λ_1, and λ_2, Eqs. (6) and (7), are

$$P(1,t) + P(2,t) + P(3,t) + P(4,t) = 1,$$
$$P(1,t) - P(4,t) = q(t), \tag{38}$$
$$P(1,t) - P(2,t) - P(3,t) + P(4,t) = r(t).$$

The three functions $\lambda_0(t), \lambda_1(t)$, and $\lambda_2(t)$ are easily obtained by Eqs. (37) and (38):

$$\lambda_0(t) = \ln 4 - \tfrac{1}{2}\ln(1-r) - \tfrac{1}{4}\ln[(1+2q+r)(1-2q+r)]$$
$$\lambda_1(t) = \tfrac{1}{2}\ln[(1-2q+r)/(1+2q+r)], \tag{39}$$
$$\lambda_2(t) = \tfrac{1}{2}\ln(1-r) - \tfrac{1}{4}\ln[(1+2q+r)(1-2q+r)].$$

As times goes to infinity, $q(t)\to 0$, $r(t)\to\eta$, the above three parameters giving the equilibrium distribution will be

$$\lambda_0^0 = \ln 4 - \tfrac{1}{2}\ln[(1-\eta)(1+\eta)],$$
$$\lambda_1^0 = 0, \tag{40}$$
$$\lambda_2^0 = \tfrac{1}{2}\ln[(1-\eta)/(1+\eta)],$$

where λ_0^0 related to the partition function Z by

$$Z = e^{\lambda_0^0} = \sum_n \exp\left(-\sum_{r=1}^{M} \lambda_r^0 A_r(n)\right). \tag{41}$$

The probability distribution functions obtained by Eqs. (37) and (38) are

$$P(1,t) = \tfrac{1}{4}(1+2q+r),$$
$$P(2,t) = P(3,t) = \tfrac{1}{4}(1-r), \tag{42}$$
$$P(4,t) = \tfrac{1}{4}(1-2q+r).$$

This result precisely coincides with the exact solution of the master equation given in Eq. (21). This coincidence is rather accidental due to the symmetry between σ_1 and σ_2 for this particular initial configuration. Still, it is remarkable that, using the knowledge of only two macroscopic observables, we could obtain the whole probability distribution functions exactly.

We next consider the case in which the observable set III is taken. In this case we let

$$A_1 = \tfrac{1}{2}(\sigma_1 + \sigma_2), \quad \langle A_1(t)\rangle = q(t), \tag{43}$$

and the corresponding Lagrange multipliers $\lambda_1(t)$. Following the similar procedure as the case of observable set III, we obtain

$$\lambda_0(t) = \ln 4 - \ln(1-q^2),$$
$$\lambda_1(t) = \ln[(1-q)/(1+q)], \tag{44}$$

and

$$P(1,t) = \tfrac{1}{4}(1+2q+q^2),$$
$$P(2,t) = P(3,t) = \tfrac{1}{4}(1-q^2), \tag{45}$$
$$P(4,t) = \tfrac{1}{4}(1-2q+q^2).$$

An interesting comparison with the exact distribution given in Eq. (42) can be made. We observe that if q^2 terms are replaced by r, the distribution functions in Eq. (45) coincide with the exact results. From Eq. (32) and Eq. (33), we notice that q^2 and r have similar time dependence

$$q^2(t) = e^{-2t}e^{2\eta t},$$
$$r(t) = (1-\eta)e^{-2t} + \eta, \tag{46}$$

if η is small. The major difference between these two functions occurs for a large t, at which $q^2\to 0$ while $r\to\eta$. We therefore conclude that the probability distribution functions obtained by the knowledge of a single observable $q(t)$ is very close to the exact distribution functions especially at the initial stage of relaxation. As the system approaches equilibrium, the agreement becomes worse, and finally it gives an incorrect distribution at equilibrium. We can very easily pinpoint the reason why it gives the wrong equilibrium distribution. The macroscopic observable we have chosen, $q(t)$, approaches zero as $t\to\infty$, and at equilibrium it provides no information at all.

Finally we would consider the other alternative, the observable set IV. In this case, we choose

$$A_1 = \sigma_1\sigma_2, \quad \langle A_1(t)\rangle = r(t), \tag{47}$$

and let the corresponding Lagrange multipliers $\lambda_1(t)$. Following the procedure used in the other cases, we obtain

$$\lambda_0(t) = \ln 4 - \tfrac{1}{2}\ln[(1-r)(1+r)],$$
$$\lambda_1(t) = \tfrac{1}{2}\ln[(1-r)/(1+r)], \tag{48}$$

and

$$P(1,t) = P(4,t) = \tfrac{1}{4}(1+r),$$
$$P(2,t) = P(3,t) = \tfrac{1}{4}(1-r). \tag{49}$$

A comparison with the exact distribution shows that the probability distribution functions in Eq. (49) are good approximation if $q(t) \ll r(t)$. This approximation is valid near the equilibrium, in which $q(t)$ vanishes. The distribution obtained by the knowledge of $r(t)$ is therefore complementary to that obtained by the knowledge of $q(t)$, each having its own range of validity.

B. Initially disaligned configuration

If we taken the initial conditions $\sigma_1 = 1, \sigma_2 = -1$ at $t = 0$, we find

$$\bar{q}(t) = \tfrac{1}{2}[q_1(t) + q_2(t)] = 0, \tag{50}$$

$$q_1(t) = e^{-(1+\eta)t}, \tag{51}$$

$$r_{12}(t) = \eta - (1+\eta)e^{-2t}. \tag{52}$$

$q_2(t)$ can be obtained explicitly from Eqs. (50) and (51). We choose the following four sets of variables as our macroscopic variables:

(I) $\bar{q}(t), q_1(t), r_{12}(t)$,

(II) $\bar{q}(t), q_1(t)$,

(III) $q_1(t), r_{12}(t)$,

(IV) $\bar{q}(t), r_{12}(t)$.

We will simply present the resulting probability distribution functions for each of the above cases, using another set of abbreviations

$$q \equiv e^{-(1+\eta)t},$$

$$r \equiv \eta - (1+\eta)e^{-2t}.$$

For observable set I,

$$\begin{aligned} P(1,t) &= P(4,t) = \tfrac{1}{4}(1+r), \\ P(2,t) &= \tfrac{1}{4}(1+2q-r), \\ P(3,t) &= \tfrac{1}{4}(1-2q-r). \end{aligned} \tag{53}$$

For observable set II,

$$\begin{aligned} P(1,t) &= P(4,t) = \tfrac{1}{4}(1-q^2), \\ P(2,t) &= \tfrac{1}{4}(1+2q+q^2), \\ P(3,t) &= \tfrac{1}{4}(1-2q+q^2). \end{aligned} \tag{54}$$

For observable set III,

$$\begin{aligned} P(1,t) &= \tfrac{1}{4}(1+r)(1+q), \\ P(2,t) &= \tfrac{1}{4}(1-r)(1+q), \\ P(3,t) &= \tfrac{1}{4}(1-r)(1-q), \\ P(4,t) &= \tfrac{1}{4}(1+r)(1-q). \end{aligned} \tag{55}$$

For observable set IV.

$$\begin{aligned} P(1,t) &= P(4,t) = \tfrac{1}{4}(1+r), \\ P(2,t) &= P(3,t) = \tfrac{1}{4}(1-r). \end{aligned} \tag{56}$$

Among these four sets of distributions, the probability distribution functions given by Eq. (53) coincide with the exact result obtained by the master equation. We notice that all three independent observables were needed to get the exact result in contrast to the case of initially aligned configuration, where only two observables were enough. All the other features are similar to the case of initially aligned configuration. The distribution obtained by the observable set IV in which $\bar{q}(t)$ and $r_{12}(t)$ are used is indentical to that which would be obtained by using a single observable $r_{12}(t)$. For this particular case the observable $\bar{q}(t)$ is redundant. The Lagrange multipliers $\lambda_0(t)$, $\lambda_1(t)$, $\lambda_2(t)$, and $\lambda_3(t)$ corresponding to the partition function, $\bar{q}(t)$, $q_1(t)$, and $r_{12}(t)$, respectively, are shown to be

$$\begin{aligned} \lambda_0(t) &= \ln 4 - \tfrac{1}{2}(1+r) \\ &\quad - \tfrac{1}{4}\ln[(1+2q-r)(1-2q-r)], \\ \lambda_1(t) &= \tfrac{1}{2}\ln[(1+2q-r)/(1-2q-r)], \\ \lambda_2(t) &= \tfrac{1}{2}\ln[(1-2q-r)/(1+2q-r)], \\ \lambda_3(t) &= \tfrac{1}{4}\ln[(1+2q-r)(1-2q-r)] - \tfrac{1}{2}\ln(1+r), \end{aligned} \tag{57}$$

which would lead to equilibrium values

$$\begin{aligned} \lambda_0^0 &= \ln 4 - \tfrac{1}{2}[(1+\eta)(1-\eta)], \\ \lambda_1^0 &= \lambda_2^0 = 0, \\ \lambda_3^0 &= \tfrac{1}{2}\ln[(1-\eta)/(1+\eta)]. \end{aligned} \tag{58}$$

Comparing Eq. (58) with Eq. (40), we find that only two nonvanishing parameters remain, each corresponding to the partition function and observable $\sigma_1\sigma_2$, respectively. We also confirm that those two parameters are independent of the initial spin configurations yielding the same equilibrium distribution. The reason why only observable $\sigma_1\sigma_2$ has nonvanishing Lagrange multiplier at equilibrium is that this is the only observable determining the equilibrium distribution. To make this point clearer, we notice that $\sigma_1\sigma_2$ is related to the energy of the system via the Hamiltonian given in Eq. (20). We can therefore relate $\langle\sigma_1\sigma_2\rangle$ to the internal energy of the system U

$$\langle\sigma_1\sigma_2\rangle = -U/J. \tag{59}$$

We expect that the corresponding Lagrange multiplier should be the inverse temperature $1/kT$ multiplied by $-J$. We readily verify that this is indeed the case, using the values of η given in Eq. (23):

$$\tfrac{1}{2}\ln[(1-\eta)/(1+\eta)] = -J/kT. \tag{60}$$

V. ENTROPY PRODUCTION

One of the fundamental concept characterizing the nonequilibrium phenomena is the entropy production discussed in Sec. II. This quantity, however, has rarely been evaluated explicitly, although there have been numerous formal discussion[4,7,8] concerning its properties. Our model presented in previous sections permits an easy evaluation of the entropy production and its time derivative as functions of time. The examination of entropy production and its derivatives evaluated explicitly and exactly might provide extra insight into the normally untractable disequilibrium processes.

For the case of initially aligned configuration we evaluate the "forces" $\mu_r(t)$, defined in Eq. (15), from Eqs. (39) and (40):

$$\mu_1(t) = \lambda_1(t) - \lambda_1^0 = \tfrac{1}{2}\ln[(1-2q+r)/(1+2q+r)], \tag{61}$$

$$\begin{aligned} \mu_2(t) &= \lambda_2(t) - \lambda_2^0 \\ &= \tfrac{1}{2}\ln(1-r) - \tfrac{1}{4}\ln[(1+2q+r)(1-2q+r)] \\ &\quad - \tfrac{1}{2}\ln[(1-\eta)/(1+\eta)]. \end{aligned}$$

The corresponding fluxes are using Eqs. (19), (34), and (35)

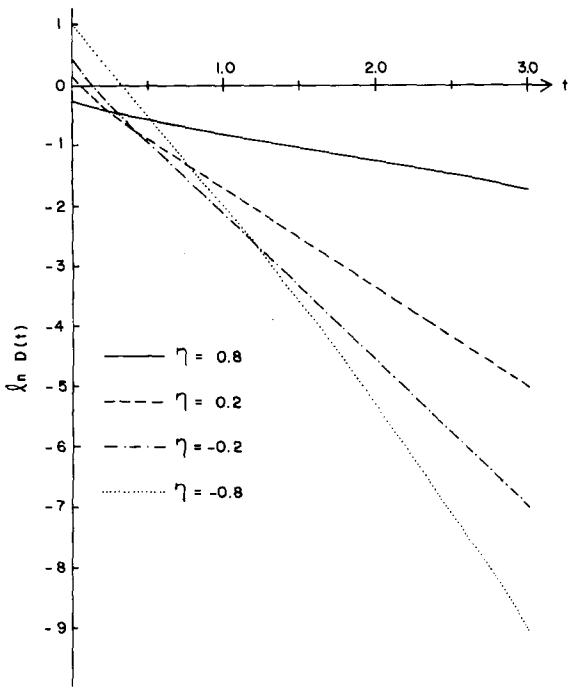

FIG. 1. Time evolution of the entropy deficiency $D(t)$ for the initially aligned configuration. The logarithm of the entropy deficiency is expressed as a function of time.

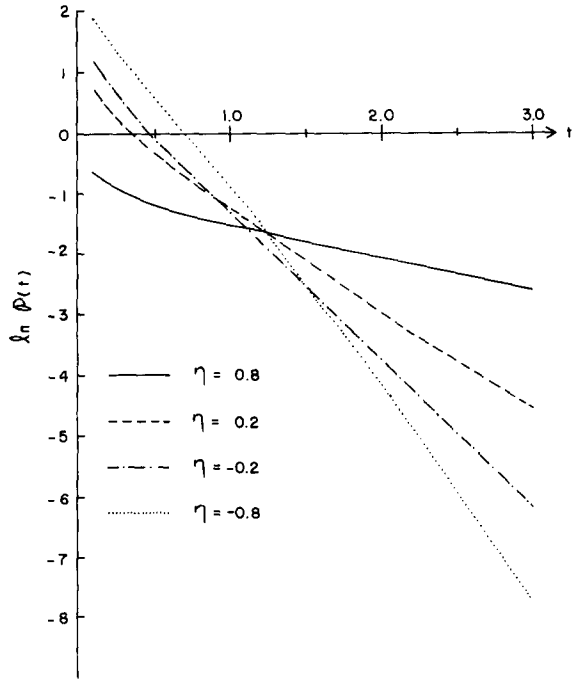

FIG. 2. Time evolution of the entropy production, $\mathscr{P}(t)$. The logarithm of $\mathscr{P}(t)$ is expressed as a function of time.

$$X_1(t) = \frac{d}{dt}q(t) = -(1-\eta)e^{-(1-\eta)t},$$
$$X_2(t) = \frac{d}{dt}r(t) = -2(1-\eta)e^{-2t}. \tag{62}$$

The entropy production $\mathscr{P}(t)$ in Eq. (18) is written explicitly in terms of t, using Eqs. (34), (61), and(62):

$$\begin{aligned}\mathscr{P}(t) = {} & \tfrac{1}{2}(1-\eta)e^{-(1-\eta)t} \\ & \times \ln\left[\frac{1+\eta+2e^{-(1-\eta)t}+(1-\eta)e^{-2t}}{1+\eta-2e^{-(1-\eta)t}+(1-\eta)e^{-2t}}\right] \\ & + \tfrac{1}{2}(1-\eta)e^{-2t} \\ & \times \left[\ln\left(\frac{1+\eta+2e^{-(1-\eta)t}+(1-\eta)e^{-2t}}{1+\eta}\right)\right. \\ & + \ln\left(\frac{1+\eta-2e^{-(1-\eta)t}+(1-\eta)e^{-2t}}{1+\eta}\right) \\ & \left. - 2\ln(1-e^{-2t})\right]. \end{aligned} \tag{63}$$

This result can also be obtained by evaluating the entropy deficiency defined in Eq. (17) and differentiating as indicated in Eq. (18). A similar expression can be obtained for the case of initially disaligned configuration. In fact, we find that the entropy production for this case is indentical to that of initially aligned configuration case shown in Eq. (63), except that the value η is replaced by $-\eta$. The time dependence of entropy deficiency and entropy production for initially aligned configuration are shown in Figs. 1 and 2, respectively. For the sake of convenience, they are plotted in log scale. It is readily verified that the entropy deficiency and entropy production are positive, which has been proved rather generally for Markov system.[2,4] As we notice from Figs. 1 and 2, $D(t)$ and $\mathscr{P}(t)$ decrease exponentially. The decay rate for positive η is slower than for negative η. This is due to the tendency of maintaining initially aligned state for positive η which describes the ferromagnetic state. The "relaxation time" may be obtained from the slopes. The slope

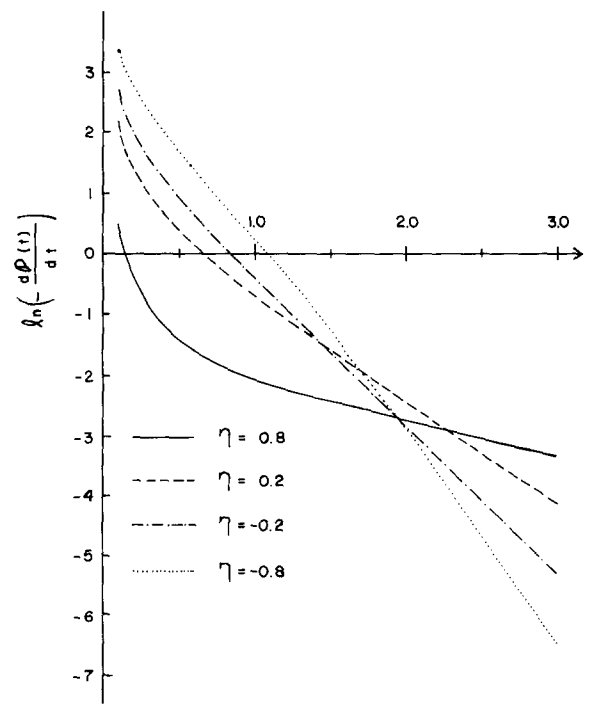

FIG. 3. Time evolution of the negative time derivative of the entropy production, $-d\mathscr{P}(t)/dt$. The logarithm of $-d\mathscr{P}(t)/dt$ is expressed as a function of time.

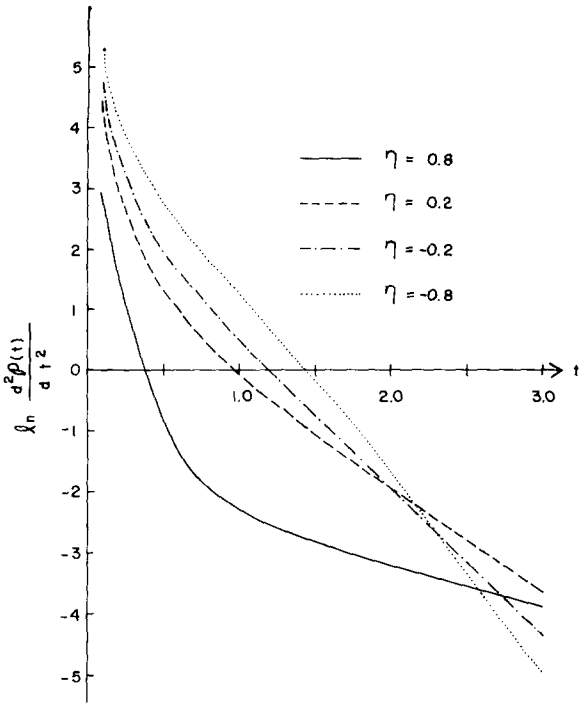

FIG. 4. Time evolution of the second time derivative of the entropy production, $d^2\mathcal{P}(t)/dt^2$. The logarithm of $d^2\mathcal{P}(t)/dt^2$ is expressed as a function of time.

for $\eta = 0.8$ is about one-third of that for $\eta = 0.2$, which implies that the "relaxation time" for $\eta = 0.8$ is three times longer than that of $\eta = 0.2$.

The negative time derivative and second time derivative of the entropy production plotted in log scale are shown in Figs. 3 and 4. The signs of these quantities are in accordance with the inequality of Prigogine which was rather generally proved by Moreau.[4]

VI. CONCLUSION

In this paper, we have rigorously tested the disequilibrium theory by applying the theory to a finite Glauber model consisting of two spins. We have obtained the conclusion that the nonequilibrium situation can be described exactly with an appropriate and sufficient number of macroscopic observables. The probability distribution functions which are obtained by the minimal information procedure then coincide exactly with the exact solution to the master equation. The nonequilibrium situation is also approximately described by the knowledge of a reduced number of macroscopic observables. In particular, if we choose the observables which do not vanish at equilibrium for the description of the system, the probability distribution functions near the equilibrium are reproduced appropriately. On the other hand, the far-from-equilibrium situation is described meaningfully only with the inclusion of some macroscopic observables which vanish at equilibrium.

As a conclusion, we expect that a disequilibrium system, of which the exact solution to the master equation is not known might be described satisfactorily if we choose a set of appropriate macroscopic observables. But the result might crucially depend on the choice of these variables.

[1] I. Procaccia, Y. Shimoni, and R.D. Levine, J. Chem. Phys. **65**, 3284 (1976).
[2] R.D. Levine, J. Chem. Phys. **65**, 3302 (1976).
[3] I. Procaccia and R.D. Levine, J. Chem. Phys. **65**, 3357 (1976).
[4] M. Moreau, J. Math. Phys. **19**, 2494 (1978).
[5] H. Spohn, J. Math. Phys. **19**, 1227 (1978).
[6] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[7] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes* (Interscience, New York, 1967).
[8] G. Nocolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems* (Wiley, New York, 1977).
[9] R.J. Glauber, J. Math. Phys. **4**, 294 (1963).

A guiding center Hamiltonian: A new approach

Robert G. Littlejohn
Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720

(Received 1 May 1979; accepted for publication 8 June 1979)

A Hamiltonian treatment of the guiding center problem is given which employs noncanonical coordinates in phase space. Separation of the unperturbed system from the perturbation is achieved by using a coordinate transformation suggested by a theorem of Darboux. As a model to illustrate the method, motion in the magnetic field $\mathbf{B} = B(x,y)\hat{z}$ is studied. Lie transforms are used to carry out the perturbation expansion.

1. INTRODUCTION

In this paper I will report on a new approach to a Hamiltonian formulation of the guiding center problem, an approach which leads to a remarkably deep insight into the formal structure of classical Hamiltonian mechanics. This insight is not new, in the sense that the natural mathematical apparatus for an abstract description of Hamiltonian mechanics is that of differential geometry, and differential geometry has been exhaustively studied by mathematicians. Nevertheless, even those mathematicians who have explicitly concerned themselves with Hamiltonian mechanics have tended to use a language and a notation which is difficult for most physicists. Among these we might mention Abraham, [1] Vinogradov and Kupershmidt, [2] and Arnold.[3] As a result, very little of the abstract point of view of Hamiltonian mechanics has found its way into the physics literature, and certainly not into the more familiar textbooks. [4–6] In addition, for most applications of interest in physics, even quite theoretical ones, a description of Hamiltonian mechanics which focuses on the differential geometry of phase space may be deemed to be unnecessarily academic and impractical. The guiding center problem appears to be an exception, however, since for this problem one is virtually compelled to employ noncanonical coordinates in phase space.

The term "the guiding center problem" refers to a certain perturbative expansion of the solution to the equations of motion of a charged particle in a given electromagnetic field. The perturbation expansion is based on an approximation, the "guiding center approximation," which may be roughly described by saying that electromagnetic effects dominate over inertial effects. This problem is of great interest and importance in plasma physics and astrophysics, and over the years various means have been devised for effecting this perturbative development. [7–14] All of these methods involve an enormous amount of algebraic manipulations, which have hindered studies into higher-order effects. For example, there still remains some controversy over certain second-order terms. This work has arisen out of an attempt to find a better way to solve this problem.

If the differential equations of motion for the guiding center probelm are written down without regard to their Hamiltonian origin, then it is straightforward but laborious to subject these equations to a systematic perturbative treatment, yielding the guiding center expansion. The required perturbation methods, which are designed for systems of ordinary differential equations with nearly periodic solutions, were largely developed by Krylov and Bogoliubov,[15] Bogoliubov and Mitropolski,[16] and Kruskal.[17] The work of Kruskal is especially significant, because he showed how the perturbative solutions relate to action integrals and adiabatic invariants in the case that the system of ordinary differential equations can be derived from a Hamiltonian.

Similar perturbative methods exist for Hamiltonian systems. These methods are older than their non-Hamiltonian counterparts, having been developed originally by Poincaré, [18] and they are the standard methods found in textbooks.[5,6,19–21] If a system can be analyzed with Hamiltonian perturbation methods, then it is much better to do so than to use non-Hamiltonian methods. The reason is that the equations of motion in Hamiltonian mechanics are derivable from a scalar function, namely the Hamiltonian, so that one can deal with a scalar instead of a vector. Similar considerations apply to coordinate transformations, which in Hamiltonian mechanics are specified by a scalar, namely the generating function of the canonical transformation. This advantage becomes greatly enhanced as one proceeds to higher and higher orders.

Unfortunately, the Hamiltonian for the guiding center problem, which will be discussed in detail in Sec. 4 below, cannot be easily analyzed by the standard methods of Poincaré. The reason is that the relation between the canonical momentum $\mathbf{p}$ and the physical variables $\mathbf{x}$ and $\mathbf{v}$ describing the motion of the particle involves the use of the magnetic vector potential $\mathbf{A}$. That is, the introduction of the vector potential is the price one must pay in order to use Hamiltonian mechanics. This in itself would not be so bad, except that in the guiding center approximation the transformation yielding $\mathbf{p}$ from $\mathbf{x}$ and $\mathbf{v}$ mixes up the ordering scheme, so that there is no clear separation between the unperturbed system and the perturbation. This difficulty is not inherent to the problem, but only to a Hamiltonian description of the problem is terms of the usual set of canonically conjugate q's and p's.

In this paper we take an approach to the guiding center problem which preserves the best features of the perturbation method of Poincaré, and yet avoids the use of the vector potential. These goals are accommplished by employing noncanonical coordinate systems in phase space. This step leads one to think more in terms of a geometrical picture of phase space dynamics, and less in terms of coordinate repre-

 0022-2488/79/122445-14$01.00

sentations with respect to canonically conjugate (q,p) pairs. One result is a heightened appreciation for the role of differential geometry in the formalism of Hamiltonian mechanics.

Sections 2 and 3 of this paper are included for the sake of establishing certain notation conventions and for the sake of completeness. Section 2 develops some of the essentials of a covariant formulation of Hamiltonian mechanics. This presentaiton is intentionally and necessarily incomplete, due to lack of space; for example, certain propositions are stated without proof. Unfortunately, there does not seem to be anything in print which covers this subject except in the abstract language of modern mathematics.

In addition, in Sec. 2 we prove a certain theorem, Theorem 1, which is not at all profound, but which seems heretofore not to have been articulated in quite the same manner, and which is crucial to our perturbation development in Sec. 5. In Sec. 3 we discuss in detail a theorem of Darboux, pertaining to the existence of canonical coordinates, which is central in our choice of coordinates in phase space.

In Sec. 4 we set up the Hamiltonian for the motion of a charged particle in the guiding center approximation. The case studied is that of a nonrelativistic particle in a static magnetic field with a high degree of symmetry, namely $\mathbf{B}(\mathbf{x}) = B(x,y)\hat{z}$. Although this is a very special case, it serves to illustrate the novel mathematical techniques described in this paper. The application of the same techniques to more realistic problems in straightforward and will be reported upon in forthcoming publications. In Sec. 4 we use a procedure suggested by the proof of Darboux's theorem to construct a certain "semicanonical" coordinate system in phase space, preparing the Hamiltonian for a standard perturbation analysis, along the lines of the method of Poincaré.

In Sec. 5 we carry out the perturbation expansion to second order in the guiding center approximation. The expansion is based on the perturbation method of Poincaré, but it differs in two significant ways. One way is that canonical transformations are expressed in terms of their Lie generators, instead of the more conventional mixed-variable generating functions. That is, we use a variant of the so-called Lie transform method, which has been pioneered by Hori,[22] Deprit,[23] Dewar,[24] and others. The second way is that a system of phase space coordinates is used which is noncanonical.

Finally, in Sec. 6 we discuss various technical aspects of the method and possible extensions and generalizations.

2. A COVARIANT FORMULATION OF HAMILTONIAN MECHANICS

In this section we outline some of the essential features of Hamiltonian mechanics in the context of an arbitrary coordinate system in phase space. To do this it is necessary to call upon the formalism of differential geometry. A relatively accessible source for a more thorough coverage of this subject is the recent textbook by Arnold. [3]

We will denote a coordinate system on phase space by the symbol $\mathbf{z}$ or z^i, representing $2N$ coordinates. N is the number of degrees of freedom of the Hamiltonian system. When these coordinates are some choice of the usual q's and p's, we will call them *canonical* coordinates, and refer to a canonical coordinate system. In this section, when we refer to canonical coordinates we will decompose the $2N$ coordinates z^i into q's and p's as follows:

$$z^i = (q_1,\ldots,q_N,p_1,\ldots,p_N). \tag{2.1}$$

Canonical coordinates are to be regarded as a special case, and unless we state the contrary, the coordinates z^i are not to be interpreted as necessarily representing a canonical coordinate system.

A convenient place to being a covariant formulation of Hamiltonian mechanics is with the Lagrange brackets. If $\mathbf{z}$ represents a set of canonical coordinates, and if $\bar{\mathbf{z}}$ represents a set of $2N$ independent functions of $\mathbf{z}$, then $\bar{\mathbf{z}}$ may be interpreted as a possibly noncanonical coordinate system in phase space. The Lagrange Bracket of the quantity $\bar{z}^i$ with the quantity $\bar{z}^j$ will be denoted by the symbol $\bar{\omega}_{ij}$, which, according to the definition, is given by

$$\bar{\omega}_{ij} = \sum_k^N \left(\frac{\partial q_k}{\partial \bar{z}^i}\frac{\partial p_k}{\partial \bar{z}^j} - \frac{\partial p_k}{\partial \bar{z}^i}\frac{\partial q_k}{\partial \bar{z}^j} \right). \tag{2.2}$$

It is convenient to introduce a certain constant, antisymmetric, orthogonal $2N \times 2N$ matrix γ, which is represented here by its partition into four $N \times N$ matrices:

$$\gamma = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{2.3}$$

In terms of the matrix γ, the Lagrange brackets $\bar{\omega}_{ij}$ can be written as follows:

$$\bar{\omega}_{ij} = \frac{\partial z^k}{\partial \bar{z}^i}\gamma_{kl}\frac{\partial z^l}{\partial \bar{z}^j}. \tag{2.4}$$

Here and throughout this section summation over repeated indices is understood.

The Poisson bracket of two phase functions f and g will be denoted by $\{f,g\}$. The Poisson brackets of the coordinates $\bar{\mathbf{z}}$ among themselves are of special importance, and we denote these quantities by $\bar{\sigma}^{ij}$. According to the definition of the Poisson bracket, we have

$$\bar{\sigma}^{ij} = \{\bar{z}^i,\bar{z}^j\} = \sum_k^N \left(\frac{\partial \bar{z}^i}{\partial q_k}\frac{\partial \bar{z}^j}{\partial p_k} - \frac{\partial \bar{z}^i}{\partial p_k}\frac{\partial \bar{z}^j}{\partial q_k} \right). \tag{2.5}$$

This can also be written in terms of the matrix γ, as follows:

$$\bar{\sigma}^{ij} = \frac{\partial \bar{z}^i}{\partial z^k}\gamma_{kl}\frac{\partial \bar{z}^j}{\partial z^l}. \tag{2.6}$$

In Eqs. (2.4) and (2.6) there may be recognized the transformation laws for the components of second-rank tensors of the covariant and contravariant types, respectively. According to this interpretation, $\bar{\omega}_{ij}$ and $\bar{\sigma}^{ij}$ are the components of two tensors with respect to the coordinate system $\bar{\mathbf{z}}$. When the coordinate system $\mathbf{z}$ is arbitrary, i.e., not necessarily canonical, or when no distinction need be made between two coordinate systems, we will drop the overbars and write simply ω_{ij} or σ^{ij} for the components of the two tensors with respect to the coordinate system $\mathbf{z}$.

The following connections between the ω tensor and the σ tensor are important. By the well-known properties of the Lagrange brackets and Poisson brackets, we have, in any coordinate system,

$$\omega_{ij}\sigma^{kj} = \delta_i^k. \tag{2.7}$$

In addition, it is easy to see that $\omega_{ij} = \sigma^{ij} = \gamma_{ij}$ if and only if the coordinate system $\mathbf{z}$ is canonical.

The ω and σ tensors can be viewed in the abstract, apart from their component representations. For the σ tensor, the relation between the two points of view is given by

$$\sigma = \sigma^{ij}\frac{\partial}{\partial z^i}\otimes\frac{\partial}{\partial z^j} = \sum_k \frac{\partial}{\partial q_k}\wedge\frac{\partial}{\partial p_k}. \tag{2.8}$$

Thus, for example, the Poisson bracket of two phase fuctions f and g can be regarded as the value of the σ tensor on the differentials of the two functions:

$$\{f,g\} = \sigma(df,dg) = \frac{\partial f}{\partial z^i}\sigma^{ij}\frac{\partial g}{\partial z^j}. \tag{2.9}$$

Likewise, the ω tensor can be regarded as a 2-form:

$$\omega = \tfrac{1}{2}\omega_{ij}dz^i\wedge dz^j = \sum_k^N dq_k\wedge dp_k. \tag{2.10}$$

The 2-form ω is *nondegenerate*, meaning

$$\det(\omega_{ij})\neq 0. \tag{2.11}$$

It is also *closed*, meaning $d\omega = 0$, or

$$\frac{\partial\omega_{ij}}{\partial z^k} + \frac{\partial\omega_{jk}}{\partial z^i} + \frac{\partial\omega_{ki}}{\partial z^j} = 0. \tag{2.12}$$

A manifold, such as Hamiltonian phase space, which is endowed with a closed, nondegenerate 2-form is said to be a *symplectic* manifold.

The fact that ω is closed is especially important. It implies and is implied by the Jacobi identity:

$$\{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} = 0. \tag{2.13}$$

We do not allow the 2-form ω to depend on time, since to do so causes the Poincaré invariants to depend on time. That is, we demand

$$\frac{\partial\omega_{ij}}{\partial t} = 0. \tag{2.14}$$

From a practical point of view, this means that most time-dependent transformations $\mathbf{z} = \mathbf{z}(\mathbf{q},\mathbf{p},t)$, taking us from a canonical coordinate system to an arbitrary system, must be excluded. Time-dependent canonical transformations are an exception, since $\omega_{ij} = \gamma_{ij} =$ constant in any canonical system. A dynamical system described by a time-dependent Hamiltonian H may be treated by the well-known procedure of taking t and $-H$ as canonically conjugate coordinates in an extended phase space of $N+1$ degrees of freedom. In this paper there will be no need to consider either time-dependent coordinate transformations or time-dependent Hamiltonians.

An important example of a noncanonical coordinate system in phase space is afforded by the dynamical system consisting of a nonrelativistic particle of mass m and charge e moving in a given, static magnetic field $\mathbf{B}(\mathbf{x})$. The usual canonical coordinates $(\mathbf{q},\mathbf{p})$ for the phase space of this system are given in terms of the particle's position $\mathbf{x}$ and velocity $\mathbf{v}$ by

$$\begin{aligned}\mathbf{q} &= \mathbf{x},\\ \mathbf{p} &= m\mathbf{v} + (e/c)\mathbf{A}(\mathbf{x}),\end{aligned} \tag{2.15}$$

where $\mathbf{A}(\mathbf{x})$ is a vector potential corresponding to the magnetic field $\mathbf{B}(\mathbf{x})$. The coordinates $(\mathbf{x},\mathbf{v})$ parametrize phase space equally as well as $(\mathbf{q},\mathbf{p})$, but they are noncanonical. Using Eq. (2.5), the components of the σ tensor with respect to this coordinate system are easily obtained:

$$\begin{aligned}&\{x_i,x_j\} = 0,\\ &\{x_i,v_j\} = -\{v_i,x_j\} = (1/m)\delta_{ij},\\ &\{v_i,v_j\} = (e/m^2c)B_{ij},\end{aligned} \tag{2.16}$$

where

$$B_{ij} = \epsilon_{ijk}B_k. \tag{2.17}$$

The components of the σ tensor can be written in matrix form, with the ordering $\mathbf{z} = (\mathbf{x},\mathbf{v})$:

$$\sigma^{ij} = \frac{1}{m}\begin{pmatrix} 0 & \vdots & I\\ \cdots & \cdots & \cdots\\ -I & \vdots & \frac{e}{mc}\mathsf{B}\end{pmatrix}. \tag{2.18}$$

Here the symbol B represents the magnetic field tensor, defined in Eq. (2.17). The components of the 2-form ω in the same coordinate system are given by

$$\omega_{ij} = m\begin{pmatrix} -\frac{e}{mc}\mathsf{B} & \vdots & I\\ \cdots & \cdots & \cdots\\ -I & \vdots & 0\end{pmatrix}. \tag{2.19}$$

Observe that the fact that ω is closed implies the Maxwell equation $\nabla\cdot\mathbf{B} = 0$.

Let us now turn our attention to Hamilton's equations of motion and their consequences. These equations are easily cast into a generally covariant form by using the Poisson bracket and Eq. (2.9). The result is

$$\frac{dz^i}{dt} = \{z^i,H\} = \sigma^{ij}\frac{\partial H}{\partial z^j}. \tag{2.20}$$

One may say that the Hamiltonian transforms as a scalar under arbitrary time-idependent coordinate transformations.

As an example of Hamilton's equations in a noncanonical coordinate system, consider the $(\mathbf{x},\mathbf{v})$ coordinate used in Eqs. (2.15)–(2.19). The Hamiltonian in the $(\mathbf{q},\mathbf{p})$ coordinates is

$$H(\mathbf{q},\mathbf{p}) = \frac{1}{2m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{q})\right)^2. \tag{2.21}$$

In the $(\mathbf{x},\mathbf{v})$ system this becomes, using Eq. (2.15),

$$H(\mathbf{x},\mathbf{v}) = \tfrac{1}{2}mv^2. \tag{2.22}$$

Then the equations of motion are

$$\begin{aligned}\frac{d}{dt}\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} &= \frac{1}{m}\begin{pmatrix} 0 & \vdots & I\\ \cdots & \cdots & \cdots\\ -I & \vdots & \frac{e}{mc}\mathsf{B}\end{pmatrix}\begin{pmatrix}\frac{\partial H}{\partial\mathbf{x}}\\ \frac{\partial H}{\partial\mathbf{v}}\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{v}\\ \frac{e}{mc}\mathbf{v}\times\mathbf{B}\end{pmatrix}.\end{aligned} \tag{2.23}$$

These are, of course, the Newton–Lorentz equations. The "nonphysical" magnetic vector potential **A** disappears from the formalism when the (**x**,**v**) coordinates are used.

Let us now return to Hamilton's equations of motion and replace the parameter t, describing the trajectories in phase space, with the nondescript parameter λ. This is done because in two applications in this paper, one in the proof of Darboux's theorem and one in the perturbation analysis of Sec. 5, the trajectories which arise from Hamilton's equations have nothing to do with the time evolution of a dynamical system. This replacement also avoids some inessential confusion over our disallowal of time-dependent coordinate transformations.

Let $\mathbf{S}(\mathbf{z}_0,\lambda)$ be the solution to Hamilton's equations which satisfies $\mathbf{z} = \mathbf{z}_0$ at $\lambda = 0$. That is $\mathbf{S}(\mathbf{z}_0,\lambda)$ satisfies

$$\frac{\partial S^i}{\partial \lambda} = \sigma^{ij}\frac{\partial H}{\partial z^j}, \tag{2.24}$$

where the right-hand side is evaluated at $\mathbf{z} = \mathbf{S}(\mathbf{z}_0,\lambda)$, and also $\mathbf{S}(\mathbf{z}_0,0) = \mathbf{z}_0$ for all $\mathbf{z}_0$. We assume the equivalent of a time-independent system, meaning that Hamilton's equations are autonomous, so that

$$\mathbf{S}(\mathbf{S}(\mathbf{z}_0,\lambda_1),\lambda_2) = \mathbf{S}(\mathbf{z}_0,\lambda_1+\lambda_2) \tag{2.25}$$

for all $\mathbf{z}_0, \lambda_1, \lambda_2$. This is an elementary result from the theory of ordinary differential equations, [25] and it gives rise to an interpretation of the solution **S** as a representation of a one-parameter group of diffeomorphisms of phase space onto itself. In view of their origin from Hamilton's equations, these diffeomorphisms are called *symplectic diffeomorphisms*, and the group is called a *Hamiltonian flow*.

Symplectic diffeomorphisms can be regarded as mappings of phase space onto itself in a manner independent of coordinate representation, or, in conjunction with a given coordinate system **z**, they can be regarded as mappings of $\mathbb{R}^{2N}$ onto itself. Of course, the underlying Hamiltonian H and *symplectic* 2-form ω are implicit. The latter point of view is more useful to us here, because it encourages us to think of symplectic diffeomorphisms as λ-dependent coordinate transformations. That is, we associate a coordinate transformation $\mathbf{z}\to\bar{\mathbf{z}}$ with $\bar{\mathbf{z}} = \mathbf{S}(\mathbf{z},\lambda)$; we will call such a coordinate transformation a *symplectic transformation*.

For the pruposes of perturbation theory it is useful to associate a symplectic transformation with a linear operator, which we denote by $T(\lambda)$. This operator acts on the vector space of phase functions and maps it into itself, according to the rule

$$(T(\lambda)f)(\mathbf{z}) = f(\mathbf{S}(\mathbf{z},\lambda)) \tag{2.26}$$

for any phase function f. That is, $Tf = f\circ\mathbf{S}$. The set $\{T(\lambda)|\lambda\in\mathbb{R}\}$ forms a linear representation of the Hamiltonian flow, and the group multiplication law, corresponding to Eq. (2.25), is

$$T(\lambda_1)T(\lambda_2) = T(\lambda_1+\lambda_2). \tag{2.27}$$

A suitable basis for the Lie algebra of the T representation of the Hamiltonian flow is the operator L, defined by

$$Lf = \{H, f\} \tag{2.28}$$

for any phase function f. With these definitions, Hamilton's equations can be written

$$\frac{d}{d\lambda}T(\lambda) = -LT(\lambda) \tag{2.29}$$

with solution

$$T(\lambda) = \exp(-\lambda L). \tag{2.30}$$

It is well known that the solutions to Hamilton's equations of motion in the usual (q,p) language give rise to canonical transformations. With respect to an arbitrary coordinate system in phase space, symplectic transformations are the proper generalization of canonical transformations, or at least the regular canonical transformations. [6] Moreover, these transformations play a privileged role among all possible transformations, in spite of the covariant formalism being pursued here, because the 2-form ω is invariant under Hamiltonian flows. This invariance can be stated in several different but equivalent ways. One way is to say that symplectic diffeomorphisms with respect to a canonical coordinate system yield canonical transformations. Another way is to state the invariance of the first Poincaré invariant, which is the integral of ω over some surface in phase space.

For our purposes we choose a third way. We consider some coordinate system **z**, with respect to which ω has components $\omega_{ij}(\mathbf{z})$, which are to be regarded as definite functions of **z**. Under an arbitrary change of coordinates $\mathbf{z}\to\bar{\mathbf{z}}$ the components of ω go into $\bar{\omega}_{ij}(\bar{\mathbf{z}})$, which we consider to be functions of the new coordinates $\bar{\mathbf{z}}$, according to the usual rule for covariant tensors:

$$\bar{\omega}_{ij}(\bar{\mathbf{z}}) = \frac{\partial z^k}{\partial \bar{z}^i}\frac{\partial z^l}{\partial \bar{z}^j}\omega_{kl}(\mathbf{z}). \tag{2.31}$$

However, if the transformation $\mathbf{z}\to\bar{\mathbf{z}}$ is a symplectic transformation, then the invariance of ω means $\bar{\omega}_{ij}(\bar{\mathbf{z}}) = \omega_{ij}(\bar{\mathbf{z}})$. Thus we have the following theorem:

Theorem 1: The functional form of the components of the 2-form ω (and hence also of the σ tensor) is invariant under symplectic transformations.

We will make use of this theorem in Sec. 5.

3. DARBOUX'S THEOREM

An axiomatic approach to Hamiltonian mechanics beings with the 2-form ω, assumed to be closed and nondegenerate, and then develops the consequences of these assumptions, such as the Jacobi identity. The approach taken in most textbooks on classical mechanics, on the other hand, is to prove theorems such as the Jacobi identity by employing a canonical coordinate system. The axiomatic approach is equivalent to the textbook approach only if it can be shown that a canonical coordinate system actually exists, i.e., a coordinate system such that $\omega_{ij} = \gamma_{ij}$. That one (and hence a whole class) does exist is a consequence of Darboux's theorem, which we shall prove in this section.

For the purposes of Darboux's theorem, it is convenient to decompose a set **z** of canonical coordinates into q's and p's in the following order:

$$\mathbf{z} = (q_1,p_1,\ldots,q_N,p_N). \tag{3.1}$$

Corresponding to this ordering, the matrix γ has the form

$$\gamma_{ij} = \begin{pmatrix} 0 & 1 & & & & & 0 \\ -1 & 0 & & & & & \\ & & 0 & 1 & & & \\ & & -1 & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & 1 \\ 0 & & & & & -1 & 0 \end{pmatrix} \tag{3.2}$$

This ordering differs from that used in Sec. 2.

We shall denote phase space by ϕ, representing a $2N$-dimensional manifold. The construction of canonical coordinates given in the proof of Darboux's theorem generally holds only locally, i.e., in some finite neighborhood of a given point. We shall, in this section, ignore all questions of the region of applicability of the construction, and speak as if it were valid for all of ϕ. With this understanding, we may state the theorem.

Theorem 2 (*Darboux's theorem*): Let there be given a closed, nondegenerate 2-form ω on ϕ and a coordinate system $\mathbf{z}$ with respect to which ω has components ω_{ij}. Then there exists a coordinate transformation $\mathbf{z}\to\bar{\mathbf{z}}$ such that the components $\bar{\omega}_{ij}$ of ω with respect to the new coordinates have the form $\bar{\omega}_{ij} = \gamma_{ij}$. Furthermore, any one of the new coordinates $\bar{z}^i$, considered as a function of the old coordinates $\mathbf{z}$, can be chosen at will.

We remark that if the original coordinate system $\mathbf{z}$ is canonical itself, then the constructive proof of Darboux's theorem gives a method of determining a canonical transformation $\mathbf{z}\to\bar{\mathbf{z}}$ in which one of the new coordinates $\bar{z}^i(\mathbf{z})$ takes on a specified form. It is in this context that Darboux's theorem will be used in Sec. 4.

Darboux's theorem is proved by induction, using the following lemma:

Lemma: Let there be given the hypotheses of Darboux's theorem. Then there exists a coordinate transformation $\mathbf{z}\to\bar{\mathbf{z}}$ such that the components $\bar{\omega}_{ij}$ of ω with respect to the new coordinates $\bar{\mathbf{z}}$ have the form

$$\bar{\omega}_{ij} = \begin{pmatrix} \Omega_{ij} & & & 0 \\ & & & \\ & & 0 & 1 \\ 0 & & & \\ & & -1 & 0 \end{pmatrix}. \tag{3.3}$$

Furthermore, any one of the new coordinates $\bar{z}^i(\mathbf{z})$ can be chosen at will.

To show how this lemma implies Darboux's theorem, we develop some simple corollaries of the lemma. To do this, it is convenient to label the new coordinates $\bar{\mathbf{z}}$ as follows:

$$\bar{\mathbf{z}} = (\mathbf{Z},q,p), \tag{3.4}$$

where the new coordinates $\mathbf{Z}$, corresponding to the Ω_{ij} block in Eq. (3.3), represent $2N-2$ functions $Z^i(\mathbf{z})$. First of all, we note that the $(2N-2)\times(2N-2)$ matrix Ω_{ij} is antisymmetric. Next, since ω is nondegenerate, we have $\det(\bar{\omega}_{ij})\neq 0$, and hence also $\det(\Omega_{ij})\neq 0$. Then, since ω is closed, we have

$$\frac{\partial\bar{\omega}_{ij}}{\partial\bar{z}^k} + \frac{\partial\bar{\omega}_{jk}}{\partial\bar{z}^i} + \frac{\partial\bar{\omega}_{ki}}{\partial\bar{z}^j} = 0. \tag{3.5}$$

If the index k in this equation is set to $2N-1$ or $2N$, corresponding to the new coordinates q or p and if neither i nor j takes on these values, then two terms vanish according to Eq. (3.3), since we have $\bar{\omega}_{jk} = \bar{\omega}_{ki} = 0$, and the remaining term gives

$$\frac{\partial\Omega_{ij}}{\partial q} = \frac{\partial\Omega_{ij}}{\partial p} = 0. \tag{3.6}$$

Hence the quantities Ω_{ij} depend only on the new coordinates $\mathbf{Z}$. When none of the indices i,j,k takes on the value $2N$–1 or $2N$, Eq. (3.5) becomes

$$\frac{\partial\Omega_{ij}}{\partial Z^k} + \frac{\partial\Omega_{jk}}{\partial Z^i} + \frac{\partial\Omega_{ki}}{\partial Z^j} = 0. \tag{3.7}$$

In Eqs. (3.6) and (3.7), the indices i,j,k run over the numbers $1,\ldots,2N-2$ corresponding to the coordinates $\mathbf{Z}$.

The result of these corollaries is that the quantities Ω_{ij} are the components with respect to the coordinate system $\mathbf{Z}$ of a certain closed, nondegenerate 2-form Ω on some manifold Φ of dimensionality $2N-2$. The manifold Φ can be identified with a submanifold of ϕ, as will be shown later. Hence on Φ the 2-form Ω satisfies the hypotheses of Darboux's theorem, and by the lemma there exists a coordinate transformation $\mathbf{Z}\to\bar{\mathbf{Z}}$, taking the components Ω_{ij} into $\bar{\Omega}_{ij}$, such that one more pair of q,p coordinates is constructed, and such that one more step toward the form of Eq. (3.2) has been taken. After N applications of the lemma, Darboux's theorem is proved.

The proof of the lemma is constructive. We will call the program for the construction of the coordinates $\bar{\mathbf{z}} = (\mathbf{Z},q,p)$ the *Darboux algorithm*.

By hypothesis, ω is nondegenerate, so $\det(\omega_{ij})\neq 0$. Therefore we can define a tensor σ with components σ^{ij} according to Eq. (2.7), and from this, a Poisson bracket according to Eq. (2.9). When we perform a coordinate transformation $\mathbf{z}\to\bar{\mathbf{z}}$, the components $\bar{\sigma}^{ij}$ of the σ tensor with respect to the new coordinates $\bar{\mathbf{z}}$ are the Poisson brackets of the new coordinates among themselves. With the definition $\bar{\mathbf{z}} = (\mathbf{Z},q,p)$, we demand the following form for these Poisson brackets:

$$\{q,p\} = 1, \tag{3.8}$$

$$\{Z^i,q\} = 0, \tag{3.9}$$

$$\{Z^i,p\} = 0, \tag{3.10}$$

$$\{Z^i,Z^j\} = \Sigma^{ij}. \tag{3.11}$$

The precise form of the quantities Σ^{ij} is immaterial, although they will automatically be the components of a $(2N-2)\times(2N-2)$, antisymmetric, invertible matrix, since the form of $\bar{\sigma}^{ij}$ is given by

$$\bar{\sigma}^{ij} = \begin{pmatrix} \Sigma^{ij} & & & 0 \\ & & & \\ & & 0 & 1 \\ 0 & & & \\ & & -1 & 0 \end{pmatrix}. \tag{3.12}$$

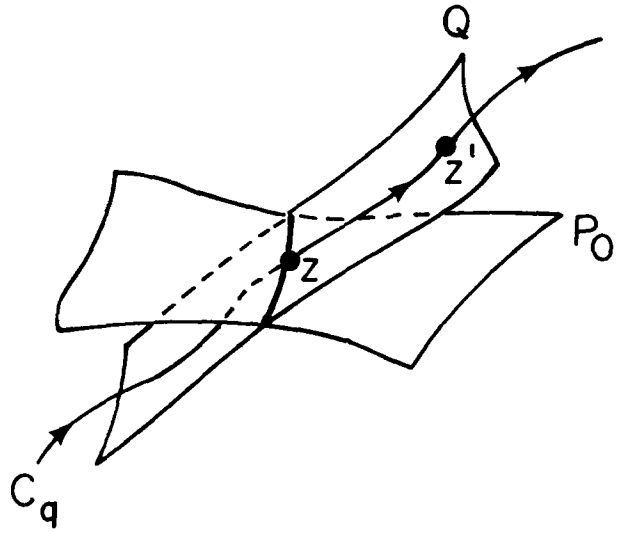

FIG. 1. The q-characteristics and the construction of the functions $p(\mathbf{z})$ and $\mathbf{Z}(\mathbf{z})$.

Clearly, Eqs. (3.8)–(3.11) are equivalent to Eq. (3.12) which in turn is equivalent to Eq. (3.3).

First we solve Eq. (3.8). We pick some function $q(\mathbf{z})$ on ϕ for one of the new coordinates; the other $2N-1$ functions, $p(\mathbf{z})$ and $Z^i(\mathbf{z})$, will then be constrained by Eqs. (3.8)–(3.10). In terms of the given function $q(\mathbf{z})$, Eq. (3.8) is a first-order, linear inhomogeneous partial differential equation for the unknown function $p(\mathbf{z})$. Such an equation always has a solution,[26] which may be found by integrating along the characteristics of the partial differential operator.

In this case the characteristics are the curves $\mathbf{z}=\mathbf{z}(\lambda)$ which are the solutions to the following set of ordinary differential equations:

$$\frac{dz^i}{d\lambda}=\{z^i,q\}. \tag{3.13}$$

These characteristics are the trajectories which result upon treating $q(\mathbf{z})$ as a Hamiltonian. Therefore, we will call them "q-characteristics." The parameter λ, which is suggestive of time, is a real number parametrizing the trajectories. It is natural to treat the operator $d/d\lambda$ as a field of tangent vectors, and to write

$$\frac{d}{d\lambda}=\sum_{ij}\sigma^{ij}\frac{\partial q}{\partial z^j}\frac{\partial}{\partial z^i}. \tag{3.14}$$

A picture of the solution $p(\mathbf{z})$ to Eq. (3.8) is useful; see Fig. 1. In this figure, Q represents a contour surface of constant q, i.e., a $(2N-1)$-dimensional manifold. Because q is constant along any q-characteristic, every q-characteristic lies in some such contour surface, such as the q-characteristic C_q in the figure. To find $p(\mathbf{z})$, we choose a $(2N-1)$-dimensional manifold P_0, cutting all the Q surfaces. P_0 is arbitrary, except that it must be nowhere tangent to any Q surface, since that would result in $dq\wedge dp=0$ and preclude the use of q and p as new coordinates. The surface P_0 is to be taken as an initial value surface for $p(\mathbf{z})$; for example, it is convenient to take $p(\mathbf{z})=0$ for $\mathbf{z}=\in P_0$. For $\mathbf{z}\notin P_0$, $p(\mathbf{z})$ is defined as the negative of the elapsed λ parameter, relative to P_0, of the q-characteristic passing through $\mathbf{z}$. From Eq. (3.14) it then follows that

$$\frac{dp}{d\lambda}=\{p,q\}=-1 \tag{3.15}$$

and Eq. (3.8) is satisfied.

Next we want to solve Eq. (3.9) for $2N-2$ functions $Z^i(\mathbf{z})$ which are independent of each other and also of q and p. Considering q as given and p and $\mathbf{Z}$ as unknowns, Eq. (3.9) is the same partial differential equation as Eq. (3.8), except that it is homogeneous. Such an equation possesses $2N-1$ independent solutions, so we seem to have one more solution than we need. Actually, we do not, because q itself satisfies the differential equation, i.e., $\{q,q\}=0$, and the remaining $2N-2$ solutions are left for the Z^i.

To construct the solutions $Z^i(\mathbf{z})$ to Eq. (3.9), observe that these functions must be constant along q-characteristics:

$$\frac{dZ^i}{d\lambda}=0. \tag{3.16}$$

The Z^i may be found by constructing a coordinate system on the surface P_0, in which q is one of the coordinates and the other $2N-2$ coordinates are Z^i. This defines $Z^i(\mathbf{z})$ for $\mathbf{z}\in P_0$. For $\mathbf{z}\notin P_0$, the values $Z^i(\mathbf{z})$ are propagated along q-characteristics so that $Z^i(\mathbf{z})=Z^i(\mathbf{z}')$ whenever $\mathbf{z}$ and $\mathbf{z}'$ are on the same q-characteristic. The result clearly satisfies Eq. (3.16), and hence also Eq. (3.9).

The functions $Z^i(\mathbf{z})$ so constructed are not unique, since any invertible transformation of the form $\overline{\mathbf{Z}}=\overline{\mathbf{Z}}(\mathbf{Z},q)$, taking $\mathbf{Z}$ into $\overline{\mathbf{Z}}$, gives a new set of solutions. Such a transformation can be regarded as a coordinate transformation on P_0.

When we turn to Eq. (3.10), we see that the Z^i must satisfy further constraints. The latitude we have in the choice of the Z^i, as mentioned in the last paragraph, is useful here, because by a proper choice of the coordinate system $(\mathbf{Z},q)$ on P_0 it is possible to satisfy Eqs. (3.9) and (3.10) simultaneously.

The characteristics of Eq. (3.10) are found by treating $p(\mathbf{z})$ as a Hamiltonian, and we will call them the "p-characteristics." They are the solutions $\mathbf{z}=\mathbf{z}(\mu)$ of the ordinary differential equations

$$\frac{dz^i}{d\mu}=\{z^i,p\}. \tag{3.17}$$

As before, we may define a tangent vector field $d/d\mu$ by

$$\frac{d}{d\mu}=\sum_{ij}\sigma^{ij}\frac{\partial p}{\partial z^j}\frac{\partial}{\partial z^i}=\{\ ,p\}. \tag{3.18}$$

The functions $Z^i(\mathbf{z})$ are to be simultaneous constants of the q-characteristics and the p-characteristics. An arbitrary pair of Hamiltonian flows does not in general possess simultaneous constants, since the diffeomorphisms belonging to the two flows do not in general commute. It may be shown, however, that two Hamiltonian flows commute if and only if the Poisson bracket of the two Hamiltonians is a constant. In the case at hand, the q-flow and the p-flow commute, since $\{q,p\}=1$.

To construct the $Z^i(\mathbf{z})$, we first select some contour surface Q_0 of $q(\mathbf{z})$, and form the $(2N-2)$-dimensional mani-

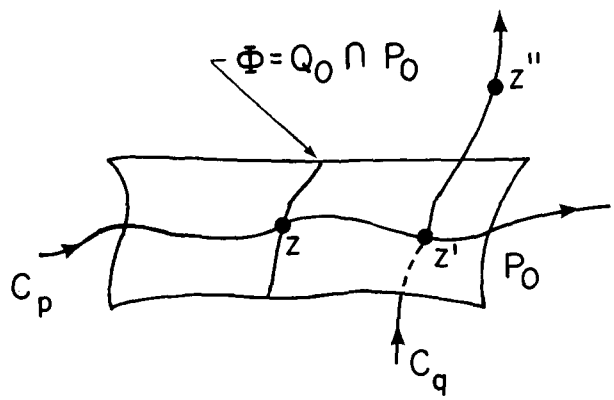

FIG. 2. The construction of the functions $\mathbf{Z}(\mathbf{z})$ as simultaneous constants of the q- and p-characteristics.

fold Φ which is the intersection of this surface with P_0, as shown in Fig. 2. The manifold Φ is the same one mentioned earlier, on which the 2-form Ω is defined. Within Φ we construct a coordinate system by arbitrarily choosing $2N-2$ independent functions $Z^i(\mathbf{z})$. Thus the $Z^i(\mathbf{z})$ are defined for $\mathbf{z}\in\Phi$. The values $Z^i(\mathbf{z})$ are then propagated along the p-characteristics passing through Φ. These characteristics lie entirely in one contour surface of p, namely P_0. Therefore the $Z^i(\mathbf{z})$ are now defined for $\mathbf{z}\in P_0$, and they are constants of the p-characteristics on this surface. The definition of the Z^i is then extended to all of ϕ by propagating along q-characteristics, as shown in Fig. 2. Thus, finally, the $Z^i(\mathbf{z})$ are defined on all of phase space, and they are constants of the q-characteristics everywhere in ϕ.

The last step is to show that the $Z^i(\mathbf{z})$ are constant of the p-characteristics, not just on P_0, but everywhere in ϕ. To do this, consider the quantities $\{Z^i,p\}$, which are known to vanish on the surface P_0. To find their values elsewhere, we compute their derivatives along the q-characteristics, using Eqs. (3.14), (3.8), and (3.9):

$$\frac{d}{d\lambda}\{Z^i,p\} = \{\{Z^i,p\},q\}$$
$$= \{\{q,p\},Z^i\} + \{\{Z^i,q\},p\} = 0 \qquad (3.19)$$

Hence the $\{Z^i,p\}$ vanish everywhere in ϕ, and Eq. (3.10) is satisfied. The Jacobi identity has entered at this point, and it is here that the fact that ω is closed, which implies the Jacobi identity, has been called upon.

This completes our proof of Darboux's theorem. Although it may be regarded as primarily of theoretical interest, we will make a practical application of it in the next section.

4. APPLICATION OF DARBOUX'S THEOREM TO THE GUIDING CENTER PROBLEM

4.1. Preliminaries

Equations (2.15) and (2.21) describe the motion of a nonrelativistic charged particle in a static magnetic field. For the purposes of this section and the next, we want to modify these equations in three steps.

The first step is to introduce a dimensionless perturbation parameter ϵ by replacing the charge e by e/ϵ. Then when the solutions to the equations of motion are developed in powers of ϵ, the result is the "guiding center approximation." Although the true solution is found in the end by setting $\epsilon = 1$, it us useful to consider ϵ to be a variable, describing a family of systems. In particular, we shall speak of the order of an expression in terms of its behavior as $\epsilon\to 0$, it being understood that the particle variables $\mathbf{x}$ and $\mathbf{v}$ and the fields $\mathbf{A}$ and $\mathbf{B}$ are to be held constant in this limiting process. For example, the gyroradius $mv_\perp c/eB$ is of order one, meaning $O(\epsilon)$, and the gyrofrequency eB/mc is of order -1, meaning $O(\epsilon^{-1})$. The physical meaning of the limit $\epsilon\to 0$ is that the particle motion is dominated by a nearly circular, rapid gyration of small gyroradius, which samples only small variations in the magnetic field during a single gyroperiod. The physical meaning of this limit is discussed in greater detail by Northrop,[8] and some of the delicate mathematical aspects of the limit are discussed by Kruskal.[9,27]

The second step is to suppress the constants e, m, and c for the sake of notational convenience. These constants are easily restored by a dimensional analysis. The resulting Hamiltonian is

$$H(\mathbf{q},\mathbf{p}) = \tfrac{1}{2}[\mathbf{p} - (1/\epsilon)\mathbf{A}(\mathbf{q})]^2 \qquad (4.1)$$

and the relation between the particle variables $(\mathbf{x},\mathbf{v})$ and the canonical coordinates $(\mathbf{q},\mathbf{p})$ is

$$\mathbf{x} = \mathbf{q},$$
$$\mathbf{v} = \mathbf{p} - (1/\epsilon)\mathbf{A}(\mathbf{q}). \qquad (4.2)$$

The third step is to restrict consideration to magnetic fields of the form $\mathbf{B}(\mathbf{x}) = B(x,y)\hat{z}$, and furthermore to consider only particle motion in the x-y plane. The problem thereby becomes two dimensional, and we write $\mathbf{x} = (x,y)$, $\mathbf{v} = (v_x,v_y)$, etc. The magnetic field can be treated as a scalar in the two-dimensional problem; we assume $B > 0$ in the region of space under consideration.

4.2. Two coordinate transformations

In this section we will subject the Hamiltonian in Eq. (4.1) to a sequence of coordinate transformations. The first is given by Eq. (4.2); it was discussed in Sec. 2 in greater detail. Under the coordinate transformation $(\mathbf{q},\mathbf{p})\to(\mathbf{x},\mathbf{v})$, the Hamiltonian becomes

$$H(\mathbf{x},\mathbf{v}) = \tfrac{1}{2}v^2. \qquad (4.3)$$

The components σ^{ij} of the σ tensor in this coordinate system can be conveniently represented by giving the formula for the Poisson bracket of two phase functions f and g:

$$\{f,g\} = \frac{\partial f}{\partial \mathbf{x}}\cdot\frac{\partial g}{\partial \mathbf{v}} - \frac{\partial f}{\partial \mathbf{v}}\cdot\frac{\partial g}{\partial \mathbf{x}} + \frac{1}{\epsilon}\mathbf{B}\cdot\left(\frac{\partial f}{\partial \mathbf{v}}\times\frac{\partial g}{\partial \mathbf{v}}\right). \qquad (4.4)$$

This is easily seen to be equivalent to Eq. (2.16). Note that ϵ appears explicitly in the Poisson bracket.

The second coordinate transformation is motivated by the form of the solution for a uniform magnetic field, which corresponds to the limit $\epsilon\to 0$. A picture of the particle motion for the case that $B(x,y)$ is uniform is shown in Fig. 3, assuming a positively charged particle. The following defini-

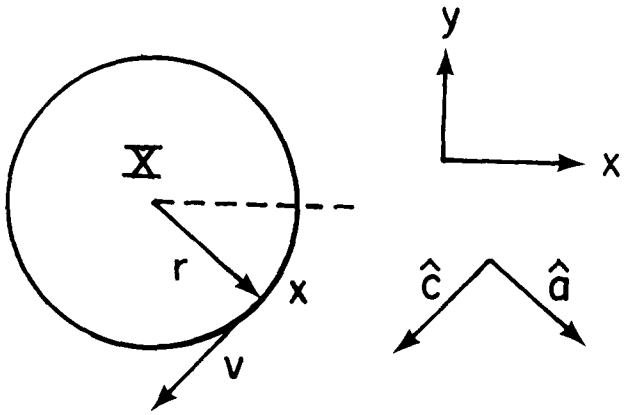

FIG. 3. Guiding center variables for a uniform magnetic field. The unit vectors $\hat{a}$, $\hat{c}$ rotate with the particle.

tions, relating to the second coordinate transformation, are valid for an arbitrary field $B(x,y)$, but their physical interpretation is most simple in the uniform case.

First we define a unit vector $\hat{\mathbf{b}}$ along the magnetic field $\mathbf{B}$. According to previous conventions, we have $\hat{\mathbf{b}} = \hat{\mathbf{z}}$. Next we define a unit vector $\hat{\mathbf{c}}$ in the direction of the particle's velocity:

$$\mathbf{v} = v\hat{\mathbf{c}}. \tag{4.5}$$

Finally, we define a unit vector $\hat{\mathbf{a}}$ by $\hat{\mathbf{a}} = \hat{\mathbf{b}} \times \hat{\mathbf{c}}$. Thus the triad $(\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{c}})$ forms a right-handed set. Note that for a uniform magnetic field $\hat{\mathbf{a}}$ is in the direction of the gyroradius vector $\mathbf{r}$, which is the displacement between the guiding center position $\mathbf{X}$ and the position of the particle $\mathbf{x}$:

$$\mathbf{x} = \mathbf{X} + \mathbf{r}. \tag{4.6}$$

In the units chosen, we have, for a uniform magnetic field,

$$\mathbf{r} = (\epsilon v/B)\hat{\mathbf{a}}. \tag{4.7}$$

Figure 3 also shows the gyrophase θ, which we define as the angle between $\hat{\mathbf{a}}$ and the x-axis, measured in a clockwise sense. Using this angle, we may state the relations between the triad $(\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\mathbf{c}})$ and $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}})$:

$$\begin{aligned} \hat{\mathbf{c}} &= -\sin\theta\hat{\mathbf{x}} - \cos\theta\hat{\mathbf{y}}, \\ \hat{\mathbf{a}} &= \cos\theta\hat{\mathbf{x}} - \sin\theta\hat{\mathbf{y}}, \\ \hat{\mathbf{b}} &= \hat{\mathbf{z}}. \end{aligned} \tag{4.8}$$

In the uniform field limit, θ evolves linearly in time with frequency B/ϵ.

We now make the coordinate transformation $(x,y,v_x,v_y) \rightarrow (x,y,\theta,v)$. The Hamiltonian keeps the form of Eq. (4.3), but the Poisson bracket changes, as indicated here by the components of the σ tensor:

$$\begin{aligned} &\{x_i,x_j\} = 0, \\ &\{\mathbf{x},v\} = \hat{\mathbf{c}}, \\ &\{\mathbf{x},\theta\} = -\hat{\mathbf{a}}/v, \\ &\{\theta,v\} = B/\epsilon v. \end{aligned} \tag{4.9}$$

4.3. The Darboux algorithm

The third coordinate transformation is not trivial, and requires some motivation. Consider a Hamiltonian $H(\mathbf{q},\mathbf{p})$. A typical strategy in Hamiltonian perturbation theory is to find a canonical transformation $(\mathbf{q},\mathbf{p}) \rightarrow (\bar{\mathbf{q}},\bar{\mathbf{p}})$ such that the new Hamiltonian K is independent of one or more (perphaps all) of the new generalized coordinates $\bar{\mathbf{q}}$. To be specific, suppose it is made independent of one new coordinate, say $\bar{q}_1$. Then none of the equations of motion for the other $2N-1$ phase coordinates depends on $\bar{q}_1$, i.e., the $\bar{q}_1$ time evolution is decoupled from the evolution of all the other phase coordinates. In addition, the conjugate momentum $\bar{p}_1$ is a constant of the motion.

It may be seen from Eq. (2.20) that such a strategy does not work so easily in the case of a noncanonical coordinate system. The Hamiltonian may be independent of one of the coordinates z^i, but it does not follow in general that some other coordinate will be a constant of the motion or that the given coordinate will decouple from the others. The reason is that consideration must be given to the components of the σ tensor, which in general depend on $\mathbf{z}$. Consider, for example, the Hamiltonian in Eq. (4.3) and the σ tensor given by Eq. (4.9). These give the following equations of motion:

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= v\hat{\mathbf{c}}, \\ \frac{dv}{dt} &= 0, \\ \frac{d\theta}{dt} &= \frac{B(\mathbf{x})}{\epsilon}. \end{aligned} \tag{4.10}$$

Thus, although $\partial H/\partial\theta = 0$, θ is not decoupled from the other variables.

It may not be necessary, however, to have a canonical coordinate system in order for the usual strategy of Hamiltonian perturbation theory to work. Consider, for example, the components of the σ tensor shown in Eq. (3.12), with respect to the coordinate system $(Z_1,..,Z_{2N-2},q,p)$. Such a coordinate system could be considered "semicanonical," because of the relations in Eqs. (3.8)–(3.10). If $\partial H/\partial q = 0$ in a coordinate system of this type, then p is a constant of the motion, and q is decoupled from the other coordinates. There is no need for the other $2N-2$ coordinates $\mathbf{Z}$ to fall into canonically conjugate pairs, and in fact it may be desirable that they not do so.

These considerations suggest that we transform from the coordinates (x,y,θ,v) to a new, semicanonical set (X,Y,θ,J), in which θ remains unchanged and J is canonically conjugate to θ, i.e., $\{\theta,J\} = 1$. The other two coordinates X and Y are to have vanishing Poisson brackets with both θ and J, but beyond that their form remains to be determined. As it turns out, these two quantities are related to a kind of generalized guiding center position.

Evidently, the coordinate transformation we desire is the result of one application of the Darboux algorithm to the coordinate set (x,y,θ,v), with θ chosen as the new generalized coordinate q, with J corresponding to p, and with $\mathbf{X} = (X,Y)$ corresponding to the $(2N-2)$-vector $\mathbf{Z}$. Actually, it is desirable to modify the form of Eqs. (3.8)–(3.10) slightly, and ask for solutions J, $\mathbf{X}$ to the set

$$\{\theta,J\} = 1/\epsilon, \tag{4.11}$$

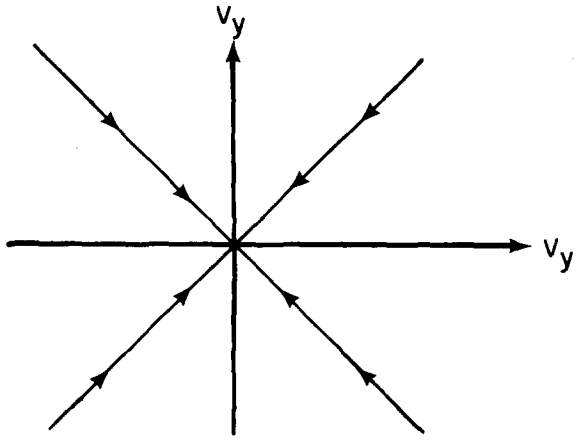

FIG. 4. The θ-characteristics converge on the surface $v = 0$.

$$\{\mathbf{X},\theta\} = 0, \tag{4.12}$$

$$\{\mathbf{X},J\} = 0. \tag{4.13}$$

The form of Eq. (4.11), which is in contrast to $\{\theta,J\} = 1$, is chosen so that the solution J will be of order zero, i.e., $O(1)$, instead of $O(\epsilon)$.

To solve these equations we will need the θ-characteristics, i.e., the trajectories which result from treating θ as a Hamiltonian. We put $d/d\lambda = \{\ ,\theta\}$ and use Eq. (4.9) to get the following differential equations for the θ-characteristics:

$$\frac{d\mathbf{x}}{d\lambda} = -\frac{\hat{\mathbf{a}}}{v}, \tag{4.14}$$

$$\frac{dv}{d\lambda} = -\frac{B(\mathbf{x})}{\epsilon v}. \tag{4.15}$$

Likewise, Eqs. (4.11) and (4.12) can be written in terms of the parameter λ:

$$\frac{dJ}{d\lambda} = -\frac{1}{\epsilon}, \tag{4.16}$$

$$\frac{d\mathbf{X}}{d\lambda} = 0. \tag{4.17}$$

To get a picture of the θ-characteristics we may examine Eq. (4.15). Since we are assuming $B > 0$, Eq. (4.15) shows that as the parameter λ increases the θ-characteristics move monotonically inward on the surfaces θ = constant toward the two-dimensional surface $v = 0$, which we shall call V_0. The projections of some of these characteristics onto the $v_x - v_y$ plane are shown in Fig. 4. It is clear that V_0 is a singular surface for the differential equations in Eqs. (4.14) and (4.15), since a single point on this surface is converged upon by a whole family of θ-characteristics, each one corresponding to a different value of θ. That this is so is not surprising, since θ has a branch point at $v = 0$. The singular nature of the θ-characteristics on this surface will cause us to make certain slight alteration in the Darboux algorithm, as it was presented in Sec. 3.

4.4. Obtaining J

To proceed, it is useful to eliminate the parameter λ from Eqs. (4.14)–(4.16) in favor of v. Since v depends monotonically on λ, this change of independent variable is permissible, and it gives

$$\frac{d\mathbf{x}}{dv} = \frac{\epsilon}{B}\hat{\mathbf{a}}, \tag{4.18}$$

$$\frac{dJ}{dv} = \frac{v}{B}. \tag{4.19}$$

Although these equations depend upon the unspecified function $B(\mathbf{x})$ and cannot, therefore, be integrated in closed form, nevertheless a perturbative solution in powers of ϵ is easily obtained. Since every θ-characteristic meets the surface V_0, the simplest initial condition to assume for the function J is $J = 0$ when $v = 0$. Then integrating Eq. (4.19) by parts and using Eq. (4.18) in an iterative manner yields the formal solution

$$J(\mathbf{x},\theta,v) = \sum_{n=0}^{\infty} \frac{(-\epsilon)^n v^{n+2}}{(n+2)!} L^n \frac{1}{B(\mathbf{x})}, \tag{4.20}$$

where L is the Lie operator defined by

$$L = \frac{1}{B(\mathbf{x})}\hat{\mathbf{a}} \cdot \frac{\partial}{\partial \mathbf{x}}. \tag{4.21}$$

The function J is our solution to Eq (4.11). Note that to lowest order it is the magnetic moment of gyration:

$$J = v^2/2B + O(\epsilon). \tag{4.22}$$

The surface V_0 corresponds, in the sense that it is the initial value surface for J, to the surface P_0 in Fig. 2 and in the discussion of the Darboux algorithm in Sec. 3. Nevertheless, it fails to correspond to P_0 in that it is two dimensional instead of three dimensional. This failure is a result of the singularity of θ on $v = 0$, and it causes V_0 to correspond, in a somewhat different sense, to the surface Φ in Fig. 2. These considerations are a warning to be careful in following the Darboux algorithm.

4.5 Obtaining X

We proceed with the construction of a simultaneous solution to Eqs. (4.12) and (4.13) as follows. First we determine the J-characteristics on V_0. We let μ be the real parameter associated with these characteristics, i.e., we put $d/d\mu = \{\ ,J\}$. In an arbitrary region of phase space the equations defining the J-characteristics are complicated, due to the complicated form of Eq. (4.20). But when $v = 0$, they simplify greatly, yielding

$$\frac{d\mathbf{x}}{d\mu} = 0, \tag{4.23}$$

$$\frac{dv}{d\mu} = 0. \tag{4.24}$$

Equation (4.24) is no surprise, because the J-characteristics must remain in a J contour surface, which is $v = 0$ by construction. As for Eq. (4.23), it tells us that the J-characteristics on V_0 are not curves at all, but rather immobile points.

Next we select a coordinate system on V_0, which is to correspond to the coordinates $\mathbf{Z}$ on Φ as described in Sec. 3, and hence also to the quantities $\mathbf{X}$ in Eqs. (4.12) and (4.13). The simplest and most obvious coordinate system is the rectangular system $\mathbf{x}$ supplied by the original problem. It is for this reason that we use the symbol $\mathbf{X}$ here instead of $\mathbf{Z}$. Therefore we define, for points on V_0,

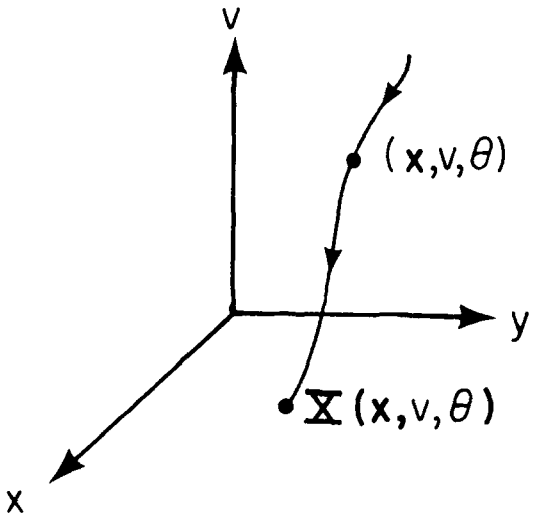

FIG. 5. Geometrical meaning of the functions $\mathbf{X}(\mathbf{x},\theta,v)$. The figure shows a θ-characteristic moving toward the surface $v=0$.

$$\mathbf{X}(\mathbf{x},v=0,\theta)=\mathbf{x}. \tag{4.25}$$

The quantities X are now propagated along J-characteristics in order to satisfy

$$\{\mathbf{X},J\}=0 \tag{4.26}$$

on V_0. But since the J-characteristics are just points, there is nothing to this step, and Eq. (4.26) is automatically satisfied on V_0.

The quantities X are next propagated along θ-characteristics to extend their definition to all of phase space. The two-dimensional surface V_0 reaches all of four-dimensional phase space by following θ-characteristics because a whole family of θ-characteristics meets any given point of V_0. The result is that the value of the function X at any given phase point $\mathbf{z}=(\mathbf{x},\theta,v)$ is found by following the θ-characteristic passing through z until it reaches $v=0$. This is shown schematically in Fig. 5. The coordinate θ has been suppressed in the figure in order to make a drawing possible. By this definition, we have

$$\{\mathbf{X},\theta\}=0 \tag{4.27}$$

everywhere in phase space.

Exactly as was done in Sec. 3, we can prove that $d/d\lambda\{\mathbf{X},J\}=0$, so that Eq. (4.26) is satisfied, not just on V_0, but everywhere in phase space. It is not at all easy to verify Eq. (4.26) directly, using the solution for J given in Eq. (4.20) and that for X given below.

At this point we find an explicit expression for the function $\mathbf{X}(\mathbf{x},\theta,v)$. This is obtained from Eq. (4.18), by means of an iterated integration by parts, exactly as Eq. (4.20) was obtained. Eq. (4.25) serves as initial conditions. The result is

$$\mathbf{X}(\mathbf{x},\theta,v)=\exp(-\epsilon vL)\mathbf{x}, \tag{4.28}$$

where the Lie operator L is defined in Eq. (4.21). It is interesting to note that when this series is carried through $O(\epsilon)$, the result is the guiding center position:

$$\mathbf{X}=\mathbf{x}-(\epsilon v/B)\hat{\mathbf{a}}+O(\epsilon^2). \tag{4.29}$$

This may be compared to Eqs. (4.6) and (4.7) for the case of the uniform magnetic field.

Our ability to express the solution X in term of a simple Lie series is probably fortuitous. For example, the analogous situation does not obtain for the guiding center problem in three dimensions. Nevertheless, some of the many properties of these series [28] will be of use to us here. For example, Eq. (4.28) may be inverted to solve for x:

$$\mathbf{x}(\mathbf{X},\theta,v)=\exp(+\epsilon vL)\mathbf{X}. \tag{4.30}$$

In this equation the Lie operator L is given by

$$L=\frac{1}{B(\mathbf{X})}\hat{\mathbf{a}}\cdot\frac{\partial}{\partial\mathbf{X}} \tag{4.31}$$

which is to be contrasted with Eq. (4.21). Lie operators are best regarded as operators which take functions into other functions, so that the independent variables in question are dummies. Therefore in what follows we shall usualy not explicity indicate the independent variables in the Lie operator itself, it being understood that they are the same as those of the operand. Equations (4.28) and (4.30) are example so this convention.

4.6. Obtaining the σ tensor

We now have an explicit form for the variable transformation $(\mathbf{x},\theta,v)\rightarrow(\mathbf{X},\theta,J)$, given by Eqs. (4.20) and (4.28). In order to make use of the new coordinate system, we need in addition the components of the σ tensor with respect to the new coordinates. Of the six independent components of the 4×4 antisymmetric component matrix σ^{ij}, five were determined by the construction of the new coordinates, as shown in Eqs. (4.11)–(4.13). The remaining component corresponds to the one independent component of the 2×2 matrix Σ^{ij}, which is shown in Eq. (3.12). This remaining component is the Poisson bracket $\{X,Y\}$, which according to Eq. (3.6) can depend only on X, i.e., not on θ or J.

Consider the Poisson bracket $\{X,Y\}$ at an arbitrary phase point $\mathbf{z}=(\mathbf{X},\theta,J)$. It is easily established that this Poisson bracket is constant along both θ- and J-characteristics, i.e., that

$$\frac{d}{d\lambda}\{X,Y\}=\frac{d}{d\mu}\{X,Y\}=0. \tag{4.32}$$

Effectively, this is an application of Poisson's theorem: The Poisson bracket of any two constants of a Hamiltonian flow is another such constant. Therefore $\{X,Y\}$ can be evaluated at any point on the θ-characteristic which passes through $\mathbf{z}=(\mathbf{X},\theta,J)$, and the result will be the same as at z itself. Clearly, the most convenient point to make such an evaluation is on V_0.

In order to find $\{X,Y\}$ on V_0 it is necessary to compute $\{X,Y\}$ in the neighborhood of V_0 and then to let $v\rightarrow0$. In this regard, it may be seen that Eq. (4.28) can be considered a power series in v as well as in ϵ. Writing this series out, and using Eq. (4.8), we have

$$\begin{aligned} X&=x-(\epsilon v/B)\cos\theta+O(v^2),\\ Y&=y+(\epsilon v/B)\sin\theta+O(v^2). \end{aligned} \tag{4.33}$$

Then a direct computation of the Poisson bracket, using Eq. (4.9), gives

$$\{X,Y\} = -\epsilon/B(\mathbf{x}) + O(v). \tag{4.34}$$

But when we let $v \to 0$, $\mathbf{x}$ becomes identical with $\mathbf{X}$, and we obtain

$$\{X,Y\} = -\epsilon/B(\mathbf{X}). \tag{4.35}$$

By the arguments above, this is valid at any point $(\mathbf{X},\theta,J)$ of phase space. As predicted, $\{X,Y\}$ depends only on $\mathbf{X}$.

Altogether, in the coordinate system (X,Y,θ,J) the components of the σ matrix are

$$\sigma^{ij} = \begin{pmatrix} 0 & -\epsilon/B(\mathbf{X}) & 0 & 0 \\ +\epsilon/B(\mathbf{X}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\epsilon \\ 0 & 0 & -1/\epsilon & 0 \end{pmatrix}. \tag{4.36}$$

That is, we can write the Poisson bracket of two functions f and g in terms of the coordinates $(\mathbf{X},\theta,J)$ as follows:

$$\{f,g\} = \frac{\epsilon}{B(\mathbf{X})}\left(\frac{\partial f}{\partial Y}\frac{\partial g}{\partial X} - \frac{\partial f}{\partial X}\frac{\partial g}{\partial Y}\right) + \frac{1}{\epsilon}\left(\frac{\partial f}{\partial \theta}\frac{\partial g}{\partial J} - \frac{\partial f}{\partial J}\frac{\partial g}{\partial \theta}\right). \tag{4.37}$$

4.7 Iterating the Darboux algorithm

At this point it is interesting to consider what would happen if another iteration of the Darboux algorithm were carried out, representing a coordinate change $(X,Y,\theta,J) \to (Q,P,\theta,J)$, which would bring the σ tensor into the form $\sigma^{ij} = \gamma_{ij}/\epsilon$. Except for the factor $1/\epsilon$, which is a minor consideration, we would then have constructed, by means of a number of noncanonical intermediaries, an overall canonical transformation $(q_x,q_y,p_x,p_y) \to (Q,P,\theta,J)$. According to the theory in Sec. 3, the new coordinates Q and P would be functions of $\mathbf{X}$ alone, and they would satisfy $\{Q,P\} = 1/\epsilon$.

The functions Q and P of $\mathbf{X}$ which are produced by a second iteration of the Darboux algorithm cannot be constructed perturbatively, as were $\mathbf{X}$ and J. Nevertheless, these functions are related in a simple manner to the well-known Euler potentials,[29] which are usually denoted by α and β:

$$Q(\mathbf{X}) = \beta(\mathbf{X})/\epsilon, \qquad P(\mathbf{X}) = \alpha(\mathbf{X})/\epsilon. \tag{4.38}$$

The functions α and β satisfy

$$\nabla\alpha \times \nabla\beta = \mathbf{B} \tag{4.39}$$

which in our two-dimensional field configuration becomes

$$B(X,Y) = \frac{\partial\alpha}{\partial X}\frac{\partial\beta}{\partial Y} - \frac{\partial\alpha}{\partial Y}\frac{\partial\beta}{\partial X}. \tag{4.40}$$

From this and Eq. (4.37) it is easy to show that $\{Q,P\} = 1/\epsilon$. Incidentally, we see that Darboux's theorem implies the existence of Euler potentials, at least for the two-dimensional field configuration considered here.

In the remainder of this paper we choose to us the coordinates $\mathbf{X}$ instead of the Euler potentials α and β, i.e., we choose to remain with the semicanonical coordinate system (X,Y,θ,J). This is done for several reasons. In the first place, what we gain by using canonical coordinates is the ability to use standard textbook formulas for Hamiltonian mechanics, while what we lose is that we must deal with Euler potentials, which are nonphysical in the same sense that the vector potential $\mathbf{A}$ is nonphysical. On the other hand, Eq. (4.37) shows that the Poisson bracket in the $(\mathbf{X},\theta,J)$ coordinate system is not excessively complicated in comparison to the usual formula for a canonical coordinate system. In the second place, when the guiding center problem is generalized to three-dimensional fields and is analyzed along the lines presented here, there results a set of four noncanonical variables, corresponding to the two variables (X,Y) given here. These four variables cannot be transformed into two canonically conjugate pairs except by using functions which are much less familiar than the Euler potentials. That is, the two-dimensional problem is a special case, in that the second application of the Darboux algorithm is solvable in terms of well-known functions. To treat the general case, it seems better to stay with noncanonical or semicanonical coordinate systems, and this we shall do also in the special two-dimensional case.

4.8. The Hamiltonian

Let us now consider the inverse of the transformation $(\mathbf{x},\theta,v) \to (\mathbf{X},\theta,J)$, which we will need in order to express the Hamiltonian in terms of the new coordinates. To begin with, we have in Eq. (4.20) the quantity J expressed as a function of $(\mathbf{x},\theta,v)$. Using Eq. (4.30), J may be expressed as a function of $(\mathbf{X},\theta,v)$. In the process of eliminating $\mathbf{x}$ in favor of $\mathbf{X}$, there results a double infinite series involving the operator L. This can be collapsed back into a single series, yielding finally

$$J(\mathbf{X},\theta,v) = \sum_{n=0}^{\infty} \frac{\epsilon^n v^{n+2}}{n!(n+2)} L^n\left(\frac{1}{B(\mathbf{X})}\right). \tag{4.41}$$

Next, we invert this series to obtain v as a function of $(\mathbf{X},\theta,J)$. Carried out through second order, this gives

$$v(\mathbf{X},\theta,J) = (2BJ)^{1/2} + \epsilon\frac{(2BJ)}{3B^2}(\hat{\mathbf{a}}\cdot\nabla B) + \epsilon^2\frac{(2BJ)^{3/2}}{72B^4} \times [9B(\hat{\mathbf{a}}\hat{\mathbf{a}}:\nabla\nabla B) - 7(\hat{\mathbf{a}}\cdot\nabla B)^2]. \tag{4.42}$$

This can then be substituted into Eq. (4.30) to obtain $\mathbf{x}$ as a function of $(\mathbf{X},\theta,J)$:

$$\mathbf{x}(\mathbf{X},\theta,J) = \mathbf{X} + \epsilon\frac{(2BJ)^{1/2}}{B}\hat{\mathbf{a}} - \epsilon^2\frac{(2BJ)}{6B^3}(\hat{\mathbf{a}}\cdot\nabla B)\hat{\mathbf{a}} + \epsilon^3\frac{(2BJ)^{3/2}}{72B^5}[-3B(\hat{\mathbf{a}}\hat{\mathbf{a}}:\nabla\nabla B) + 5(\hat{\mathbf{a}}\cdot\nabla B)^2]\hat{\mathbf{a}}. \tag{4.43}$$

In Eqs. (4.42) and (4.43), B means $B(\mathbf{X})$ and ∇ means $\partial/\partial\mathbf{X}$. These two formulas give the desired inverse transformation, $(\mathbf{X},\theta,J) \to (\mathbf{x},\theta,v)$.

Finally, we can use Eqs. (4.3) and (4.42) to find the Hamiltonian in the $(\mathbf{X},\theta,J)$ coordinate system. The result is

$$H(\mathbf{X},\theta,J) = BJ + \epsilon\frac{(2BJ)^{3/2}}{3B^2}(\hat{\mathbf{a}}\cdot\nabla B)$$

$$+\epsilon^2\frac{(2BJ)^2}{24B^4}[3B(\hat{\mathbf{a}}\hat{\mathbf{a}}:\nabla\nabla B)-(\hat{\mathbf{a}}\cdot\nabla B)^2]$$
$$+O(\epsilon^3). \tag{4.44}$$

In the next section we will follow the usual strategy of Hamiltonian perturbation theory in order to find a transformation which will make H independent of θ. The result will be a Hamiltonian for the guiding center motion.

5. THE GUIDING CENTER HAMILTONIAN

In this section the Hamiltonian in Eq. (4.44) is subjected to a near-identity coordinate transformation of the form $(X,\theta,J)\rightarrow(\bar{X},\bar{\theta},\bar{J})$ such that three criteria are fulfilled. First, the new Hamiltonian is to be independent of $\bar{\theta}$. Second, the transformation is to be free of secular terms. And third, the new coordinates are to be semicanonical in the same sense that the old ones are, so that $\bar{J}$ will be a constant of the motion (the generalized magnetic moment) and so that the time evolution of $\bar{\theta}$ will decouple from that of the other phase coordinates. The first two criteria are standard in Hamiltonian perturbation theory for nearly periodic systems; the third is a novel element, arising from our use of noncanonical coordinates in phase space.

We are not looking for canonical transformations, in the usual sense, because our coordinate system is noncanonical. However, on the strength of Theorem 1, we do want to use symplectic transformations, since these will cause the third criterion to be fulfilled. Although these coordinate transformations are very much like canonical transformations, being in a sense canonical transformations expressed in noncanonical coordinates, it is nevertheless awkward to express them in terms of the usual mixed variable generating functions. Instead, we express these symplectic transformations in terms of a set of Lie generators, following the theory outlined in Sec. 2. Our procedure has been thoroughly discussed by Dragt and Finn.[30]

Consider a sequence $w_1,w_2,\cdots$ of time-independent phase functions, and associated operators $L_1,L_2,\cdots$ which are defined on analogy to Eq. (2.28):

$$L_n f=\epsilon\{w_n,f\} \tag{5.1}$$

for any phase function f. The factor ϵ has been introduced into this definition because the Poisson bracket given in Eq. (4.37) has a term which is $O(\epsilon^{-1})$.

Next, each of these functions is used to generate a symplectic transformation, according to the formula

$$T_n=\exp(-\epsilon^n L_n/n). \tag{5.2}$$

The factor $1/n$ is included in order to make the resulting formulas follow as closely as possible the conventions of Cary.[31] Finally, a symplectic transformation T is constructed by multiplying together the T_n:

$$T=\cdots T_3T_2T_1, \tag{5.3}$$
$$T^{-1}=T_1^{-1}T_2^{-1}T_3^{-1}\cdots. \tag{5.4}$$

These operators are expanded as power series in ϵ by multiplying together the exponential series associated with Eq. (5.2). To obtain the correct ordering in powers of ϵ it is necessary to take account of the fact that the operators L_n consist of a $O(1)$ part and an $O(\epsilon^2)$ part, according to Eq. (4.37). Therefore we define two more series of operators, as follows:

$$M_n f=\frac{\partial w_n}{\partial\theta}\frac{\partial f}{\partial J}-\frac{\partial w_n}{\partial J}\frac{\partial f}{\partial\theta} \tag{5.5}$$

and

$$N_{n+2}f=\frac{1}{B}\left(\frac{\partial w_n}{\partial Y}\frac{\partial f}{\partial X}-\frac{\partial w_n}{\partial X}\frac{\partial f}{\partial Y}\right) \tag{5.6}$$

so that

$$L_n=M_n+\epsilon^2N_{n+2}. \tag{5.7}$$

When the operators T and T^{-1} are expressed in terms of the M and N operators, the results are, through third order in ϵ,

$$T=I-\epsilon M_1+\tfrac{1}{2}\epsilon^2(-M_2+M_1^2)+\tfrac{1}{6}\epsilon^3$$
$$\times(-2M_3-6N_3-M_1^3+3M_2M_1)+O(\epsilon^4), \tag{5.8}$$

$$T^{-1}=I+\epsilon M_1+\tfrac{1}{2}\epsilon^2(M_2+M_1^2)+\tfrac{1}{6}\epsilon^3(2M_3+6N_3$$
$$+M_1^3+3M_1M_2)+O(\epsilon^4). \tag{5.9}$$

In terms of the coordinates $\mathbf{z}=(\mathbf{X},\theta,J)$ and $\bar{\mathbf{z}}=(\bar{\mathbf{X}},\bar{\theta},\bar{J})$, we may say, somewhat loosely,

$$\bar{\mathbf{z}}=T\mathbf{z}, \tag{5.10}$$
$$\mathbf{z}=T^{-1}\bar{\mathbf{z}}. \tag{5.11}$$

As was noted in Sec. 4, the independent variables of the Lie operators M_n and N_n which appear in the expansion of T are the same as those of the operand.

When the symplectic transformation T is applied to the Hamiltonian H, there results a new Hamiltonian K, according to

$$K=T^{-1}H. \tag{5.12}$$

In this equation we expand both K and H in powers of ϵ:

$$H=\sum_{n=0}^{\infty}\epsilon^n H_n, \tag{5.13}$$

$$K=\sum_{n=0}^{\infty}\epsilon^n K_n. \tag{5.14}$$

Then using Eq. (5.9) and collecting terms gives a hierarchy of equations, which through second order can be expressed as follows:

$$0=K_0-H_0, \tag{5.15}$$
$$M_1H_0=K_1-H_1, \tag{5.16}$$
$$M_2H_0=2(K_2-H_2)-M_1(H_1+K_1). \tag{5.17}$$

These equations are written in this form because they are to be regarded as partial differential equations for the w_n, which specify the transformation T. To see this, note that

$$M_nH_0=B\frac{\partial w_n}{\partial\theta}. \tag{5.18}$$

The perturbation expansion is carried out by selecting the w_n, order by order, so that K is independent of θ, and so that the w_n contain only purely oscillatory terms in θ. The resulting w_n are

$$w_1=[(2BJ)^{3/2}/3B^3](\hat{\mathbf{c}}\cdot\nabla B), \tag{5.19}$$
$$w_2=[(2BJ)^2/24B^5]\hat{\mathbf{a}}\hat{\mathbf{c}}:(3B\nabla\nabla B-\nabla B\nabla B). \tag{5.20}$$

The new Hamiltonian K, which we may justifiably call the guiding center Hamiltonian, is given by

$$K(\bar{\mathbf{X}},\bar{J}) = B\bar{J} + \epsilon^2(\bar{J}^2/4B^2)[B\nabla^2 B - 3(\nabla B)^2] + O(\epsilon^3), \tag{5.21}$$

where B means $B(\bar{\mathbf{X}})$ and where ∇ means $\partial/\partial\bar{\mathbf{X}}$.

The equations of motion resulting from K are immediate; the effect of the ϵ ordering of the Poisson bracket should be noted.

$$\frac{d\bar{\mathbf{X}}}{dt} = \frac{\epsilon\hat{b}}{B}\times\left\{\bar{J}\nabla B + \epsilon^2\frac{\bar{J}^2}{4}\nabla\left[\frac{\nabla^2 B}{B} - \frac{3(\nabla B)^2}{B^2}\right]\right\} + O(\epsilon^5), \tag{5.22}$$

$$\frac{d\bar{\theta}}{dt} = \frac{B}{\epsilon} + \epsilon\frac{\bar{J}}{2B^2}[B\nabla^2 B - 3(\nabla B)^2] + O(\epsilon^3), \tag{5.23}$$

$$\frac{d\bar{J}}{dt} = 0. \tag{5.24}$$

The first term of Eq. (5.22) is the so-called "grad B drift."

Finally, the relation (5.10) can be written out, connecting $\mathbf{z}$ and $\bar{\mathbf{z}}$. This gives

$$\bar{\mathbf{X}} = \mathbf{X} + \frac{\epsilon^3}{B}\hat{b}\times\nabla\left[\frac{(2BJ)^{3/2}}{3B^3}(\hat{\mathbf{c}}\cdot\nabla B)\right] + O(\epsilon^4), \tag{5.25}$$

$$\bar{\theta} = \theta + \epsilon\frac{(2BJ)^{1/2}}{B^2}(\hat{\mathbf{c}}\cdot\nabla B) + \epsilon^2\frac{(2BJ)}{12B^4}\hat{\mathbf{a}}\hat{\mathbf{c}}:(3B\nabla\nabla B - 5\nabla B\nabla B) + O(\epsilon^3), \tag{5.26}$$

$$\bar{J} = J + \epsilon\frac{(2BJ)^{3/2}}{3B^3}(\hat{\mathbf{a}}\cdot\nabla B) + \epsilon^2\frac{(2BJ)^2}{48B^5}[(7\hat{\mathbf{a}}\hat{\mathbf{a}} + 9\hat{\mathbf{c}}\hat{\mathbf{c}}) :\nabla B\nabla B + 3B(\hat{\mathbf{a}}\hat{\mathbf{a}} - \hat{\mathbf{c}}\hat{\mathbf{c}}):\nabla\nabla B] + O(\epsilon^3). \tag{5.27}$$

In all cases these formulas have been carried out to the highest order which is consistent with the knowledge of only w_1 and w_2.

By combining Eqs. (5.25)–(5.27) with (4.20) and (4.28) the variables $(\bar{\mathbf{X}},\bar{\theta},\bar{J})$ can be expressed in terms of $(\mathbf{x},\mathbf{v})$. We remark that although the convergence of the series in Eqs.(5.25)–(5.27) is questionable, the convergence of the series in Eqs. (4.20) and (4.28) is easy to establish for sufficiently small values of ϵ and for $1/B$ a real analytic function of $\mathbf{x}$. The practical utility of perturbation series may not be lost even if the series are divergent.

6. DISCUSSION AND CONCLUSIONS

The use of the transformation given in Eqs. (4.20) and (4.28), which we may call the Darboux transformation, is the most unusual element in the approach taken in this paper to a perturbation problem. There is nothing new, however, in the function which this transformation serves. The Darboux transformation fulfills the purpose of isolating the unperturbed system from the perturbation, and it is exactly the difficulty of achieving this separation that has made previous Hamiltonian treatments of guiding center motion so nonstandard in appearance and awkward in execution. In addition, the Darboux transformation yields a set of variables which are natural to the unperturbed system, since to lowest order $\mathbf{X}$ and J are constants of the motion and θ evolves linearly in time. The importance of these two goals—the isolation of the unperturbed system and the choice of an appropriate set of coordinates for the unperturbed system—has been made very clear, on the basis of an invariant, geometrical picture of phase space orbits, in a seminal paper by Kruskal [17] on nearly periodic systems. These goals are common to both Hamiltonian and non-Hamiltonian systems, and the Darboux transformation forms a kind of bridge between a Hamiltonian and a non-Himiltonian treatment of the guiding center problem.

In textbook problems on pertubation theory the unperturbed system is separated from the perturbation at the outset, and hence the separation, as a task in itself, is hardly recognized. In a non-Hamiltonian treatment of the guiding center problem it is nearly trivial to achieve this separation, as has been shown by Bogoliubov and Mitropolski. [16] It was on the basis of this non-Hamiltonian separation that the angle θ was chosen as a new coordinate in the construction of the Darboux transformation in Sec. 4, and this choice caused the desired separation in the Hamiltonian treatment as well.

Likewise, the choice of appropriate variables for the unperturbed system is often nearly unconscious in textbook examples. In Hamiltonian systems, this choice can be formalized by saying that one must solve the Hamilton–Jacobi equation for the unperturbed system before proceeding with a perturbation treatment, although often the required solution is obvious. In our example, the Darboux transformation automatically provides us with a set of coordinates appropriate to the unperturbed system, because the canonically conjugate variables θ and J are effectively action-angle variables for the unperturbed system.

The construction of the Darboux transformation, as it was given in Sec. 4, is not unique, in the sense that the selection of any phase function which differs from θ by terms of order ϵ or higher would satisfy the two goals discussed above equally as well as θ itself. The only reason for choosing θ is that it has simple dependence on $(\mathbf{x},\mathbf{v})$. Indeed, if $\bar{\theta}$, given by Eq. (5.26), were chosen, then not only would the unperturbed system separate from the perturbation, but also the entire Hamiltonian would decouple from $\bar{\theta}$. This consideration raises the possibility that the construction of the Darboux transformation in Sec. 4 and the perturbation treatment in Sec. 5 could be merged, although I have not yet investigated this question.

In this paper a Hamiltonian treatment of the guiding center problem has been achieved at the expense of the construction of the Darboux transformation. It may well be asked if the result is worth the price. There are several reasons to believe the answer is yes.

In the first place, even if the results are carried to lowest order, giving only the classic, well-known "drifts," the method provides, nonetheless, a Hamiltonian treatment of these lowest-order results within the framework of a systematic ordering scheme.

Second, the method seems to give the shortest avenue to higher-order results, in terms of the labor involved, although this may best be judged by those who have used other methods. The perturbation treatment in Sec. 5 is no worse than any standard Hamiltonian perturbation treatment, and

enormously better than a non-Hamiltonian treatment. The Darboux transformation itself is perturbative, i.e., it is a power series in ϵ instead of a transformation in closed form, but it is based on a secular perturbation treatment which is quite simple. On balance, it seems that a simple secular perturbation expansion plus a standard Hamiltonian perturbation expansion is much less laborious than a non-Himiltonian expansion.

Third, a simple Hamiltonian treatment of the guiding center problem opens the door to the addition of other perturbations, such as electromagnetic waves, and to the study of, for example, the effects of these on adiabatic invariants. some results along these lines have already been achieved by Grebogi, Kaufman, and Littlejohn. [32]

Fourth, successive iterations of the Darboux algorithm give a simple means of exploring the other adiabatic invariants of guiding center motion, such as the longitudinal invariant and the flux invariant. [8]

Fifth, since the dynamics of statistical ensembles of charge particles in the Vlasov approximation can be described in Hamiltonian terms, the guiding center Hamiltonian can be used to treat nonuniform magnetic fields in a plasma, a case of great practical importance. The possible applications of a guiding center Hamiltonian to kinetic theory are too numerous to mention.

Several extensions of the results of the present paper have already been completed and will be reported upon in forthcoming publications. Two-dimensional, fully electromagnetic fields have been treated, as well as three-dimensional magnetostatic fields. The results are promising, and work is beginning on three-dimensional electromagnetic fields and relativistic treatments, as well as on applications in other directions.

ACKNOWLEDGMENTS

I would like to thank A.N. Kaufman for his patient encouragement of this work, and to acknowledge several helpful discussion with R.K. Sachs and J.R. Cary. In addition, critical readings of the manuscript by C. Grebogi, J.A. Rome, A. Ogawa, and A. Mine are greatly appreciated.

This work was supported by the U.S. Department of Energy under contract W-7405-ENG-48.

[1]R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
[2]A.M. Vinogradov and B.A. Kupershmidt, Russ. Math. Surveys **32**, 177 (1977).
[3]V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer–Verlag, New York, 1978).
[4]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, Mass. 1950).
[5]E. Saletan and A. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971).
[6]E.C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Approach* (Wiley, New York, 1974).
[7]H. Alfvén, *Cosmical Electrodynamics* (Clarendon, Oxford, 1950).
[8]T.G. Northrop, *The Adiabatic Motion of Charged Particles* (Interscience, New York, 1963).
[9]M.D. Kruskal, "Elementary Orbit and Drift Theory," in *Plasma Physics* (International Atomic Energy Agency, Vienna, 1965).
[10]A. Baños, J. Plasma Phys. **1**, 305 (1967).
[11]P. Martín, "Second Order Drift Velocities," preprint, Departamento Física, Universidad Simón Bolívar, Caracas, Venezuela, 1977.
[12]C.S. Gardner, Phys. Rev. **115**, 791 (1959).
[13]T.G. Northrop and J.A. Rome, Phys. Fluids **21**, 384 (1978).
[14]D.P. Stern, "Hamiltonian Formulation of Guiding Center Motion," Goddard Space Flight Center Report No. X-641-71-56.
[15]N. Krylov and N. Bogoliubov, *Introduction to Non-Linear Mechanics* (Princeton University, Princeton, N.J., 1947).
[16]N.N. Bogoliubov and Y.A. Mitropolski, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (Gordon and Breach, New York, 1961).
[17]M.D. Kruskal, J. Math. Phys. **3**, 806 (1962).
[18]H. Poincare, *Les Methods Nouvelles de la Mechanique Celeste*, 1892 (Reprinted by Dover, New York, 1957).
[19]D. ter Haar, *Elements of Hamiltonian Dynamics* (North-Holland, Amsterdam, 1964).
[20]Y. Hagihara, *Celestial Mechanics* (MIT, Cambridge, Mass., 1972).
[21]A.H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).
[22]G. Hori, Publ. Astron. Soc. Japan **18**, 287 (1966).
[23]A. Deprit, Cel. Mech. **1**, 12 (1969).
[24]R.L. Dewar, J. Phys. A **9**, 2043 (1976).
[25]V.I. Arnold, *Ordinary Differential Equations* (MIT, Cambridge, Mass., 1978).
[26]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953).
[27]M.D. Kruskal, "Asymptotology," *Mathematical Models in Physical Sciences* (Prentice-Hall, Englewood Cliffs, 1963).
[28]W. Gröbner, *Die Lie-Reihen und Ihre Anwendungen* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1967).
[29]D.P. Stern, Am. J. Phys. **38**, 494 (1970).
[30]A.J. Dragt and J.M. Finn, J. Math. Phys. **17**, 2215 (1976).
[31]J.R. Cary, "Lie Transforms and Their Use in Hamiltonian Perturbation Theory," Lawrence Berkeley Laboratory Report No. LBL-6350, 1978.
[32]C. Grebogi, A.N. Kaufman, and R.G. Littlejohn, "Ponderomotive Effects Due to an Electromagnetic Wave in a Nonumiform Magnetic Field," Lawrence Berkeley Laboratory Report No. LBL-8916, 1979.

World-line invariance in predictive mechanics[a)]

Angel Salas
Department of Physics, Temple University, Philadelphia, Pennsylvania 19122 [b)]

(Received 22 September 1978; accepted for publication 9 August 1979)

The Currie–Hill conditions for the relativistic world-line invariance of a Newtonian-like dynamical system of interacting particles are generalized to cover the case of invariance under any given finite-dimensional continuous group of transformations of space–time. Necessary and locally sufficient conditions are obtained both in the general case and in the important particular case when the group of transformations includes time translations as a subgroup.

I. INTRODUCTION AND SUMMARY

In 1960 Havas and Plebański[1] showed that, for a system of interacting particles, Newtonian causality, also called predictivity,[2] is not incompatible with special relativity, as had been generally thought until them. Subsequently, there has been a gradual revival of interest in the study of Newtonian-like dynamical systems as the basis for a possible alternative to the field theories which are normally associated with the fundamental interactions of physics.[3] An important result in the study of dynamical systems with Newtonian causality is that of Currie[4] and Hill[5] giving a system of first-order partial differential equations that the force functions have to satisfy in order to have relativistic world-line invariance. Later Bel[6] proved that those conditions were also sufficient. In this paper we generalize that result, obtaining necessary and locally sufficient conditions for the world-line invariance under any finite-dimensional continuous group of transformations of space–time, for a dynamical system with Newtonian causality.[7]

In Sec. II some indications on the notation used are given. In Sec. III the single-time and multiple-time formulations of the equations of motion for a system of N classical structureless point particles with Newtonian causality are studied in detail as a preparation for the main results to be obtained later. No essential original contribution is to be found here.

In Sec. IV we start with a given realization of a Lie group G_r on space–time $\mathbb{R}^{D+1}$ (Sec. IV.A) and from it we develop the naturally induced realizations on $\mathbb{R}^{2D+1}$ (Sec. IV.B) and on $(\mathbb{R}^{2D+1})^N$ (Sec. IV.C). In Sec. IV.C the concept of invariance of the dynamical system under G_r is defined, and it is proved that a necessary and (locally) sufficient set of conditions for that invariance is $[\mathbf{H},\mathbf{N}_{(L)}]=0$. The vector field $\mathbf{H}$ is the generator of the group of transformations $(t_a, x_a^i, v_a^i)\mapsto(t_a+T, x_a'^i, v_a'^i)$, where $x_a'^i$ and $v_a'^i$ are the (components of the) positions and velocities at $t_a'=t_a+T$ along the dynamical trajectories going through x_a^i at t_a with velocities v_a^i, and $\mathbf{N}_{(L)}$ are the generators of the constant-time transformations $(t_a, x_a^i, v_a^i)\mapsto(t_a, \tilde{x}_a^i, \tilde{v}_a^i)$, which give the same images of the dynamically possible trajectories as the given realization of G_r when applied to each particle independently of the others. [See Fig. 1, for two particles; (Γ_1,Γ_2) is a pair of dynamically possible trajectories, and $f_A(\Gamma_1)$ and $f_A(\Gamma_2)$ are the images of those trajectories under the transformation f_A of the group.] In Sec. IV.D the main results of this paper are derived. Theorem 1 states that the invariance conditions $[\mathbf{H},\mathbf{N}_{(L)}]=0$ on $(\mathbb{R}^{2D+1})^N$ are equivalent to $[\mathbf{h},\boldsymbol{\eta}_{(L)}]=0$, where $\mathbf{h}$ and $\boldsymbol{\eta}_{(L)}$ are the vector fields on $\mathbb{R}^{2DN+1}$ corresponding to $\mathbf{H}$ and $\mathbf{N}_{(L)}$ on $(\mathbb{R}^{2D+1})^N$, i.e., referring to Fig. 2, $\mathbf{h}$ and $\boldsymbol{\eta}_{(L)}$ are the generators of $(t, x_a^i, v_a^i)\mapsto(t+T, x_a'^i, v_a'^i)$ and $(t, x_a^i, v_a^i)\mapsto(t, \tilde{x}_a^i, \tilde{x}_a^i)$, respectively. Theorem 2 applies in the important particular case when the group of transformations includes time translations as a subgroup. Basically, it expresses the necessary and (locally) sufficient conditions for the invariance of the system under G_r in terms of commutation relations between the $\boldsymbol{\eta}_{(L)}$'s.

Finally, in the Conclusion, Sec. V, the main results of this paper are discussed.

II. NOTATION

Throughout this paper the following indices are used:

(i) $a, b, \ldots = 1, \ldots, N$ (labeling the particle);

(ii) $i, j, \ldots = 1, \ldots, D$, where D is the dimensionality of space (normally $D=3$);

(iii) $\alpha, \beta, \ldots = 0, 1, \ldots, D$;

(iv) $I, J, \ldots = 1, \ldots, 2D$;

(v) $\lambda, \mu, \ldots = 0, 1, \ldots, 2D$;

(vi) $L, M, \ldots$ label the parameters of the Lie group G_r.

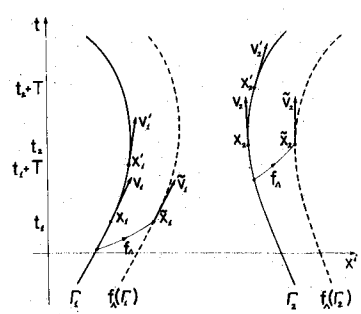

FIG. 1. Transformations generated by the vector fields $\mathbf{H}$ and $\mathbf{N}$ (space–time diagram for two particles and dimensionality of space $D=1$).

[a)]This paper is based on the first part of the Ph.D. thesis of the author (Temple University, 1978).

[b)]Present address: Departament de Física Teórica, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain.

 0022-2488/79/122459-08$01.00

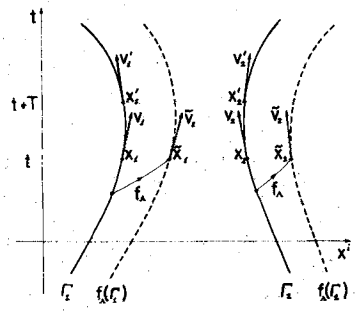

FIG. 2. Transformations generated by the vector fields $\mathbf{h}$ and $\boldsymbol{\eta}$.

The summation convention is used for all kinds of indices except those labeling the particles (a, b, ...). We assume $a' \neq a$.

$$(x^\alpha) \equiv (x^0, x^i) \equiv (t, x^i)\,; \tag{2.1}$$

$$(X^I) \equiv (x^i, v^i)\,; \tag{2.2}$$

$$(X^\lambda) \equiv (X^0, X^I) \equiv (x^\alpha, v^i) \equiv (t, x^i, v^i)\,; \tag{2.3}$$

$$(X^I_a) \equiv (x^i_a, v^i_a)\,; \tag{2.4}$$

$$X \equiv (X^\lambda_a) \equiv (X^0_a, X^I_a) \equiv (t_a, x^i_a, v^i_a)\,; \tag{2.5}$$

$$y \equiv (t, X^I_a) \equiv (t, x^i_a, v^i_a)\,. \tag{2.6}$$

Since the treatment in this paper is strictly local,[7] we can identify the space–time manifold with $\mathbb{R}^{D+1} = \{(x^\alpha)\}$. Similarly, we consider $\mathbb{R}^{2D+1} = \{(X^\lambda)\}$, $(\mathbb{R}^{2D+1})^N = \{X\}$, and $\mathbb{R}^{2DN+1} = \{y\}$.

The parameters of the Lie group G_r are designated Λ^L, and $(\Lambda^L) \equiv \Lambda$. The unit element is Λ_0. The structure constants of G_r are C^N_{LM}. Partial derivatives are sometimes denoted by commas:

$$f^\alpha{}_{,\beta} \equiv \frac{\partial f^\alpha}{\partial x^\beta}, \quad \xi^\alpha_{(L),\beta} \equiv \frac{\partial \xi^\alpha_{(L)}}{\partial x^\beta}, \quad \xi^\alpha_{(L),\beta\gamma} \equiv \frac{\partial^2 \xi^\alpha_{(L)}}{\partial x^\gamma \partial x^\beta}, \tag{2.7}$$

$$\xi^\alpha_{(L)a,\beta} \equiv \frac{\partial \xi^\alpha_{(L)a}}{\partial x^\beta_a}, \quad \xi^\alpha_{(L)a,\beta\gamma} \equiv \frac{\partial^2 \xi^\alpha_{(L)a}}{\partial x^\gamma_a \partial x^\beta_a}.$$

III. NEWTONIAN CAUSALITY

A. Single-time formulation

We consider a system of N point particles with coordinates x^i_a. The equations of motion are said to have Newtonian causality if the specification of the positions and velocities at any given time t_0 determines the trajectories completely and uniquely.[1] In other words, there is Newtonian causality if the dynamically possible trajectories are given by

$$x^i_a = \varphi^i_a(y_0; t)\,, \tag{3.1}$$

where φ^i_a are functions such that, for all $y_0 \equiv (t_0, x^i_{0a}, v^i_{0a})$,

$$\varphi^i_a(y_0; t_0) = x^i_{0a}\,, \qquad \dot{\varphi}^i_a(y_0; t_0) = v^i_{0a}\,, \tag{3.2}$$

with

$$\dot{\varphi}^i_a(y_0; t) \equiv \frac{\partial}{\partial t}\varphi^i_a(y_0; t)\,.$$

From Eq. (3.1) there follows

$$\frac{dx^i_a}{dt} = \dot{\varphi}^i_a(y_0; t)\,, \tag{3.3}$$

$$\frac{d^2x^i_a}{dt^2} = \ddot{\varphi}^i_a(y_0; t)\,. \tag{3.4}$$

Taking $t = t_0$ in Eq. (3.4) we get

$$\left\{\frac{d^2x^i_a}{dt^2}\right\}_{t_0} = \ddot{\varphi}^i_a(y_0; t_0) \equiv \mu^i_a(y_0)\,.$$

Because of Eqs. (3.2) and (3.3) this is equivalent to $\{d^2x^i_a/dt^2\}_{t_0} = \{\mu^i_a(t, x^i_b, dx^i_b/dt)\}_{t_0}$, and since t_0 is arbitrary, we finally obtain

$$\frac{d^2x^i_a}{dt^2} = \mu^i_a\left(t, x^i_b, \frac{dx^i_b}{dt}\right). \tag{3.5}$$

Thus, Eqs. (3.1) imply Eq. (3.5). Conversely, given the second-order system of ordinary differential equations (3.5), there exists a unique solution (3.1) satisfying Eq. (3.2). (However, this solution may be only local, i.e., valid for t in a certain neighborhood of t_0.[7]) Therefore the equations of motion have Newtonian causality if and only if they are equivalent to a second-order system of ordinary differential equations of the form (3.5).

In order to have a system of first-order differential equations instead of second-order ones it is convenient to introduce the new variables v^i_a with dynamical trajectories

$$v^i_a = \dot{\varphi}^i_a(y_0; t)\,. \tag{3.6}$$

Then the system (3.5) is equivalent to the first-order system [cf. Eq. (2.4)]

$$\frac{dX^I_a}{dt} = h^I_a(t, X^J_b)\,, \tag{3.7}$$

with $(h^I_a) = (v^i_a, \mu^i_a)$. In terms of the $(2DN+1)$-dimensional variable y [cf. Eq. (2.6)] the system (3.7) is equivalent to

$$\frac{dy}{dt} = \mathbf{h}(y)\,, \tag{3.8}$$

where $\mathbf{h}$ is the vector field on $\mathbb{R}^{2DN+1}$,

$$\mathbf{h} \equiv \frac{\partial}{\partial t} + \sum_a h^I_a \frac{\partial}{\partial X^I_a}. \tag{3.9}$$

The solution of Eq. (3.8) through y_0 is given in terms of the exponential mapping by

$$y = e^{(t-t_0)\mathbf{h}}(y_0)\,, \tag{3.10}$$

i.e.,

$$e^{(t-t_0)\mathbf{h}}(y_0) = (t, \varphi^i_a(y_0; t), \dot{\varphi}^i_a(y_0; t))\,. \tag{3.11}$$

B. Multiple-time formulation[8]

So far the initial data were all given at the same time t_0. We now consider a set of initial data $\bar{X}^I_a$ given at times $\bar{X}^0_a \equiv \bar{t}_a$, possibly different for each particle. Consider the set of equations

$$\bar{X}^I_a = \Phi^I_a(X^J_{0b}; \bar{t}_a)\,, \tag{3.12}$$

where

$$\Phi^I_a(X^J_{0b}; t) \equiv [e^{t\mathbf{h}}(t_0 = 0, X^J_{0b})]^I_a\,. \tag{3.13}$$

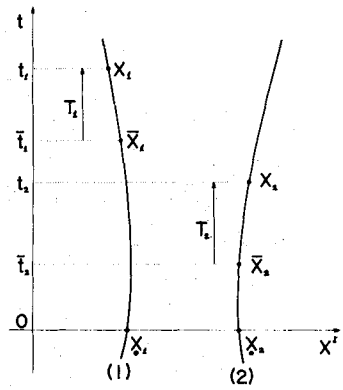

FIG. 3. Initial data at times $\bar{t}_a$ (schematic diagram for two particles on $\mathbb{R}^{2D+1}$, with $\mathbb{R}^{2D}$ reduced to $\mathbf{R}$).

Since

$$\left\{\frac{\partial \Phi^I_a}{\partial X^J_{0b}}\right\}_{t=0} = \delta^b_a \delta^I_J ,$$

and thus the Jacobian $\det(\{\partial \Phi^I_a / \partial X^J_{0b}\}_{t=0})$ is not zero, the implicit-function theorem tells us that, for values of the $\bar{t}_a$'s small enough (in absolute value), there exist unique functions $\Psi^I_a(\bar{X})$ such that

$$\Phi^I_a[\Psi^J_b(\bar{X}); \bar{t}_a] \equiv \bar{X}^I_a . \tag{3.14}$$

Thus, Eq. (3.12) is equivalent to

$$X^J_{0b} = \Psi^J_b(\bar{X}) , \tag{3.15}$$

and we have

$$X^I_a = \Phi^I_a(X^J_{0b}; t_a) = \Phi^I_a[\Psi^J_b(\bar{X}); t_a] . \tag{3.16}$$

It is convenient to measure times on each trajectory from the corresponding initial time t_a. To this end we define

$$T_a \equiv t_a - \bar{t}_a \equiv X^0_a - \bar{X}^0_a . \tag{3.17}$$

We also define the functions Φ^0_a as follows:

$$\Phi^0_a(X^J_{0b}; t) \equiv t , \tag{3.18}$$

and then Eq. (3.16) is equivalent to

$$X^\lambda_a = \Omega^\lambda_a(\bar{X}; T_a) , \tag{3.19}$$

with

$$\Omega^\lambda_a(\bar{X}; T_a) \equiv \Phi^\lambda_a[\Psi^J_b(\bar{X}); \bar{X}^0_a + T_a] . \tag{3.20}$$

In particular, we have

$$\Omega^0_a(\bar{X}; T_a) = \bar{X}^0_a + T_a , \tag{3.21}$$

$$\Omega^\lambda_a(\bar{X}; 0) = \bar{X}^\lambda_a . \tag{3.22}$$

Lemma: The Ω's constitute a realization of an N-parameter additive group (the T's being the parameters):

$$\Omega^\lambda_a[\Omega^\mu_b(\bar{X}; T_b); T'_a] = \Omega^\lambda_a(\bar{X}; T_a + T'_a) . \tag{3.23}$$

Proof: First we notice that

$$\Psi^I_a[\Omega^\mu_b(\bar{X}; T_b)] = \Psi^I_a(\bar{X}) , \tag{3.24}$$

as follows from

$$\Phi^\lambda_a[\Psi^J_b(\bar{X}); \Omega^0_a(\bar{X}; T_a)] = \Omega^\lambda_a(\bar{X}; T_a)$$

[which is an immediate consequence of Eqs. (3.20) and (3.21)] and the definition of Ψ^J_b [Eq. (3.14)]. Now, making the substitutions $\bar{X}^\mu_b \to \Omega^\mu_b(\bar{X}; T_b)$ and $T_a \to T'_a$ in Eq. (3.20) and using Eqs. (3.24), (3.21), and (3.20) in the right hand side of the resulting expression, we obtain Eq. (3.23). This completes the proof.

From the theory of continuous groups of transformations[9] it is known that the necessary and sufficient condition for Eq. (3.23) to be satisfied is that there exist N mutually commuting vector fields $\mathbf{H}_{(b)}$ on $(\mathbb{R}^{2D+1})^N$ such that Eq. (3.19) is the solution of the system of partial differential equations

$$\frac{\partial X}{\partial T_b} = \mathbf{H}_{(b)}(X) , \tag{3.25}$$

satisfying the initial conditions (3.22). Since Ω^λ_a is not a function of $T'_a (a' \neq a)$, the vector fields $\mathbf{H}_{(b)}$ have to be of the form

$$\mathbf{H}_{(b)} = H^\lambda_b \frac{\partial}{\partial X^\lambda_b} \tag{3.26}$$

(no sum over b, see Sec. II), and taking this into account Eq. (3.25) is equivalent to

$$\frac{\partial X^\lambda_a}{\partial T_b} = \delta^b_a H^\lambda_a(X) . \tag{3.27}$$

Once it is known that the system (3.7) can be written in the multiple-time form (3.27), it is easy to obtain the expression for H^λ_a. We get

$$H^0_a(X) = 1 , \tag{3.28a}$$

$$H^I_a(X) = h^I_a\{\Phi^J_b[\Psi^K_c(X); X^0_a], X^0_a\} . \tag{3.28b}$$

The commutability conditions for the generators $\mathbf{H}_{(b)}$,

$$[\mathbf{H}_{(a)}, \mathbf{H}_{(b)}] = 0 , \tag{3.29}$$

which are nothing other than the conditions for complete integrability of system (3.25), are equivalent to

$$\mathscr{L}_{\mathbf{H}_{(a')}} H^\lambda_a \equiv H^\mu_{a'} \frac{\partial H^\lambda_a}{\partial X^\mu_{a'}} = 0 \quad (a' \neq a) \tag{3.30}$$

because of Eq. (3.26). In view of this commutability the solution of Eq. (3.25) with initial value $\bar{X}$ when $T_b = 0$ can be written as

$$\begin{aligned} X &= e^{\sum_b T_b \mathbf{H}_{(b)}}(\bar{X}) \\ &= e^{T_1 \mathbf{H}_{(1)}} \cdots e^{T_N \mathbf{H}_{(N)}}(\bar{X}) \end{aligned} \tag{3.31}$$

[the exponential factors in the right hand side of Eq. (3.31) can be taken in any order]. Thus,

$$\Omega^\lambda_a(\bar{X}; T_a) = \left[e^{\sum_b T_b \mathbf{H}_{(b)}}(\bar{X})\right]^\lambda_a , \tag{3.32}$$

with the values of the $T_{a'}$'s $(a' \neq a)$ in the right hand side of the equation arbitrary. In particular, we can take $T_{a'} = T_a = T$; then we have, with $\Omega \equiv (\Omega^\lambda_a)$,

$$\Omega(\bar{X}; T) = e^{T\mathbf{H}}(\bar{X}) , \tag{3.33}$$

where $\mathbf{H}$ is the vector field

$$\mathbf{H} = \sum_b \mathbf{H}_{(b)} = \sum_b H^\lambda_b \frac{\partial}{\partial X^\lambda_b} . \tag{3.34}$$

What Eq. (3.33) says is that $X = \Omega(\bar{X}; T)$ is the solution of

$$\frac{dX}{dT} = \mathbf{H}(X), \tag{3.35}$$

with the initial condition (3.22). Thus, the system of partial differential equations (3.25) is equivalent to the system of ordinary differential equations (3.35) because of the commutability of the $\mathbf{H}_{(b)}$'s, and both Eqs. (3.25) and (3.35) are equivalent to the system (3.8) on $\mathbb{R}^{2DN+1}$. More explicitly, this means that the general solutions of the three systems (3.8), (3.25), and (3.35) define the same family of N-tuples of geometrical trajectories in $\mathbb{R}^{2D+1}$ in either of the following two cases:

(i) the vector field $\mathbf{h} = (\partial/\partial t) + \Sigma_b h^I_a(\partial/\partial X^I_a)$ is given (with arbitrary h^I_a), the $\mathbf{H}_{(b)}$'s are defined by Eq. (3.26) with (3.28), and $\mathbf{H}$ is defined by Eq. (3.34);

(ii) the vector field $\mathbf{H} = \Sigma_b H^\lambda_b(\partial/\partial X^\lambda_b)$ is given, satisfying Eqs. (3.28a) and (3.30) (and otherwise arbitrary), the $\mathbf{H}_{(b)}$'s are defined by Eq. (3.26), and $\mathbf{h}$ is defined by Eq. (3.9) with

$$h^I_a(t, X^J_b) \equiv H^I_a(X^0_b = t, X^J_b). \tag{3.36}$$

Notice that when we start with the single-time formulation [case (i)] the commutability conditions (3.30) are a consequence of the definition of H^λ_a, while when we start directly with the multiple-time formulation [case (ii)] those commutability conditions have to be required independently.

In the preceding paragraph it was not assumed that the first-order system (3.8) was equivalent to a second-order system (3.5). If this is the case we have to require $h^i_a = v^i_a$ in case (i), from which there follows $H^i_a = v^i_a$, and *vice versa* in case (ii). Then the identity of N-tuples of geometrical trajectories in $\mathbb{R}^{2D+1}$ is equivalent to the identity of their projection on space–time $\mathbb{R}^{D+1}$.

IV. WORLD-LINE INVARIANCE

In this section we establish necessary and sufficient conditions for the invariance of a dynamical systems with Newtonian causality under a group of transformations of space–time. The main result of this paper, obtained in Sec. IV.D, is the derivation of a set of such necessary and (locally) sufficient conditions expressed in terms of vector fields on $\mathbb{R}^{2DN+1}$. This result is not obvious because the action of the group is given on $\mathbb{R}^{D+1}$ (space–time) rather than on $\mathbb{R}^{2DN+1}$ directly.[10]

For theorem 2 at the end of the section it will be required of the continuous group of transformations to include time translations, but apart from that the group is arbitrary.

A. Realization of G_r on $\mathbb{R}^{D+1}$

We assume that a realization of an r-parameter Lie group G_r on $\mathbb{R}^{D+1}$ is given. Let

$$x'^\alpha = f^\alpha(x^\beta; \Lambda^L) \equiv f^\alpha_\Lambda(x^\beta) \tag{4.1}$$

define this realization (for one particular parametrization of G_r, Λ^L being the parameters). The generators of the realization are

$$\xi_{(L)} \equiv \xi^\alpha_{(L)} \frac{\partial}{\partial x^\alpha} \equiv \left\{ \frac{\partial f^\alpha}{\partial \Lambda^L} \right\}_{\Lambda_0} \frac{\partial}{\partial x^\alpha}. \tag{4.2}$$

B. Realization of G_r on $\mathbb{R}^{2D+1}$

We want to extend the realization (4.1) of G_r on $\mathbb{R}^{2D+1}$ to obtain a realization on $\mathbb{R}^{2D+1}$ which is equivalent to the former for every trajectory $x^i = \varphi^i(t)$ (nor necessarily a solution of any equations of motion) if we take $v^i = \dot{\varphi}^i(t)$. We have

$$X'^\lambda = F^\lambda(X^\mu; \Lambda^L) \equiv F^\lambda_\Lambda(X^\mu), \tag{4.3}$$

with

$$F^\alpha(X^\mu; \Lambda^L) \equiv f^\alpha(x^\beta; \Lambda^L),$$

$$F^{D+i}(X^\mu; \Lambda^L) \equiv \frac{df^i/dt}{df^0/dt} \equiv \frac{f^i_{,0} + v^j f^i_{,j}}{f^0_{,0} + v^j f^0_{,j}}. \tag{4.4}$$

The natural components of the generators of this realization, $\Xi^\lambda_{(L)} \equiv \{\partial F^\lambda/\partial \Lambda^L\}_{\Lambda_0}$, are

$$\Xi^\alpha_{(L)} = \xi^\alpha_{(L)}, \tag{4.5}$$

$$\Xi^{D+1}_{(L)} = \xi^i_{(L),0} + v^j \xi^i_{(L),j} - v^i(\xi^0_{(L),0} + v^j \xi^0_{(L),j}).$$

C. Realization of G_r on $(\mathbb{R}^{2D+1})^N$

The generalization of the realization (4.3) of G_r on $\mathbb{R}^{2D+1}$ to a realization on $(\mathbb{R}^{2D+1})^N$ is straightforward. We assume that for each particle we have a realization of the form (4.3), independently of the other particles. Thus, we get

$$X'^\lambda_a = F^\lambda(X^\mu_a; \Lambda^L) \equiv F^\lambda_a(X; \Lambda), \tag{4.6}$$

i.e.,

$$X' = F(X; \Lambda) \equiv F_\Lambda(X). \tag{4.7}$$

For the generators we have

$$\Xi_{(L)} = \sum_a \Xi^\lambda_{(L)a} \frac{\partial}{\partial X^\lambda_a} \equiv \sum_a \left\{ \frac{\partial F^\lambda_a}{\partial \Lambda^L} \right\}_{\Lambda_0} \frac{\partial}{\partial X^\lambda_a} \tag{4.8}$$

and, from Eq. (4.6),

$$\Xi^\lambda_{(L)a}(X) = \Xi^\lambda_{(L)}(X^\mu_a). \tag{4.9}$$

We can now define the concept of world-line invariance under (the given realization of) G_r. We say that the equations of motion have world-line invariance under G_r, or simply that they are invariant under G_r, if, when we transform each dynamically possible N-tuple of geometrical trajectories point by point using Eq. (4.6), each new N-tuple of geometrical trajectories obtained is also dynamically possible (i.e., satisfies the equations of motion), for any value of Λ. Analytically, this means that there are N functions T'_a of T, Λ^L, and X^λ_b such that

$$\{F_\Lambda \circ e^{T\mathbf{H}}(X)\}^\lambda_a = \{e^{T'_a(T;\Lambda;X)\mathbf{H}} \circ F_\Lambda(X)\}^\lambda_a \tag{4.10}$$

are identically satisfied. It can be proved[11] that a necessary and sufficient condition for it is that there exist $N^2 r$ scalars on $(\mathbb{R}^{2D+1})^N$, c^b_{aL}, such that

$$[\Xi_{(L)}, \mathbf{H}_{(a)}] = \sum_b c^b_{aL} \mathbf{H}_{(b)}; \tag{4.11}$$

because of Eq. (3.26) this is equivalent to

$$[\Xi_{(L)}, \mathbf{H}_{(a)}]^{\lambda}_{a} = c_{aL} H^{\lambda}_{a}, \tag{4.12}$$

with $c^{b}_{aL} = \delta^{b}_{a} c_{aL}$. Taking $\lambda = 0$ in Eq. (4.12), we obtain $c_{aL} = -\mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)a}$; inserting this in Eq. (4.12) again and using

$$[\mathbf{H}, \mathbf{H}_{(b)}] = 0 \tag{4.13}$$

[which is a consequence of Eqs. (3.29) and (3.34)], we obtain

$$[\Xi_{(L)}, \mathbf{H}] = \left[\sum_{b} \Xi^{0}_{(L)b} \mathbf{H}_{(b)}, \mathbf{H} \right],$$

i.e.,

$$[\mathbf{H}, \mathbf{N}_{(L)}] = 0, \tag{4.14}$$

with

$$\mathbf{N}_{(L)} \equiv \Xi_{(L)} - \sum_{b} \Xi^{0}_{(L)b} \mathbf{H}_{(b)}. \tag{4.15}$$

Therefore, Eqs. (4.14) are necessary and (locally) sufficient conditions for the invariance of the equations of motion under G_r.

The geometrical meaning of $\mathbf{N}_{(L)}$ (which has $N^{0}_{(L)a} = 0$) was illustrated in Sec. I by means of Fig. 1. In the next subsection we will be more explicit about $\boldsymbol{\eta}_{(L)}$, but what will be said there about the interpretation of $\boldsymbol{\eta}_{(L)}$ applies, *mutatis mutandis*, to $\mathbf{N}_{(L)}$.

D. Realization of G_r on $\mathbb{R}^{2DN+1}$

The result just derived may be new, but the previous results presented in this paper were not really original, although some aspects of their presentation may have been. Actually, Sec. III and the foregoing part of this Sec. IV were only preparation for the new theorems to be proved presently, which are a generalization of previous results of Currie,[4] Hill,[5] and Bel.[6,2]

Theorem 1: A necessary and (locally) sufficient condition for the invariance of equations of motion (3.8) under G_r is that the following set of commutation relations between vectors on $\mathbb{R}^{2DN+1}$ be identically satisfied:

$$[\mathbf{h}, \boldsymbol{\eta}_{(L)}] = 0, \tag{4.16}$$

where

$$\eta^{0}_{(L)}(t, X^{J}_{b}) \equiv 0,$$
$$\eta^{I}_{(L)a}(t, X^{J}_{b}) \equiv N^{I}_{(L)a}(X^{0}_{b} = t, X^{J}_{b}) \equiv \Xi^{I}_{(L)}(t, X^{J}_{a}) - \Xi^{0}_{(L)}(t, X^{J}_{a}) h^{I}_{a}(t, X^{J}_{b}), \tag{4.17}$$

and, as seen before [Eqs. (3.9) and (3.36)],

$$h^{0} = 1,$$
$$h^{I}_{a}(t, X^{J}_{b}) \equiv H^{I}_{a}(X^{0}_{b} = t, X^{J}_{b}). \tag{4.18}$$

Proof: We have to prove that Eq. (4.14) $\Longleftrightarrow$ Eq. (4.16), i.e.,

$$\{[\mathbf{H}, \mathbf{N}_{(L)}] = 0\} \Longleftrightarrow \{[\mathbf{h}, \boldsymbol{\eta}_{(L)}] = 0\}, \tag{4.19}$$

where the left-hand side of the double implication is on $(\mathbb{R}^{2D+1})^{N}$ and the right-hand side is on $\mathbb{R}^{2DN+1}$.

That Eq. (4.14) $\Rightarrow$ Eq. (4.16) is trivial: Take $X^{0}_{b} = t$ (for all b) in Eq. (4.14). To prove that Eq. (4.16) $\Rightarrow$ Eq. (4.14) we define

$$\mathbf{V}_{(L)} \equiv [\mathbf{H}, \mathbf{N}_{(L)}] \quad [\text{on } (\mathbb{R}^{2D+1})^{N}] \tag{4.20}$$

(which implies $V^{0}_{(L)a} = 0$), and we have to prove that $\mathbf{V}_{(L)} = 0$ if Eq. (4.16) is satisfied. First we notice that

$$\mathbf{V}_{(L)} = \left[\mathbf{H}, \Xi_{(L)} - \sum_{b} \Xi^{0}_{(L)b} \mathbf{H}_{(b)}\right] = [\mathbf{H}, \Xi_{(L)}] - \sum_{b} (\mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)b}) \mathbf{H}_{(b)} \tag{4.21}$$

because of Eq. (4.13). Now we have

$$\begin{aligned}
&[\mathbf{H}_{(c)}, \mathbf{V}_{(L)}] \\
&= \left[\mathbf{H}_{(c)}, [\mathbf{H}, \Xi_{(L)}] - \sum_{b} (\mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)b}) \mathbf{H}_{(b)}\right] \\
&= [\mathbf{H}_{(c)}, [\mathbf{H}, \Xi_{(L)}]] - \sum_{b} (\mathscr{L}_{\mathbf{H}_{(c)}} \mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)b}) \mathbf{H}_{(b)} \\
&= [\mathbf{H}, [\mathbf{H}_{(c)}, \Xi_{(L)}]] - \sum_{b} (\mathscr{L}_{\mathbf{H}} \mathscr{L}_{\mathbf{H}_{(c)}} \Xi^{0}_{(L)b}) \mathbf{H}_{(b)} \\
&= [\mathbf{H}, [\mathbf{H}_{(c)}, \Xi_{(L)}]] - (\mathscr{L}_{\mathbf{H}} \mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)c}) \mathbf{H}_{(c)}
\end{aligned}$$

[in going from the second to the third line use has been made of the Jacobi identities and Eq. (4.13); for the last line we have used

$$\mathscr{L}_{\mathbf{H}(c)} \Xi^{0}_{(L)b} = \delta^{c}_{b} \mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)c}].$$

For $c = a' \neq a$ this gives

$$[\mathbf{H}_{(a')}, \mathbf{V}_{(L)}]^{\lambda}_{a} = [\mathbf{H}, [\mathbf{H}_{(a')}, \Xi_{(L)}]]^{\lambda}_{a} = -[\mathbf{H}, \Xi_{(L)}]^{\mu}_{a'} \frac{\partial H^{\lambda}_{a}}{\partial X^{\mu}_{a'}}, \tag{4.22}$$

using

$$[\mathbf{H}_{(a')}, \Xi_{(L)}]^{\mu}_{b} = \delta^{b}_{a'} [\mathbf{H}, \Xi_{(L)}]^{\mu}_{b}.$$

From Eq. (4.21) we have

$$[\mathbf{H}, \Xi_{(L)}]^{\mu}_{a'} = V^{\mu}_{(L)a'} + (\mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)a'}) H^{\mu}_{a'}$$

and when we use this in Eq. (4.22) we get

$$[\mathbf{H}_{(a')}, \mathbf{V}_{(L)}]^{\lambda}_{a} = -V^{\mu}_{(L)a'} \frac{\partial H^{\lambda}_{a}}{\partial X^{\mu}_{a'}} - (\mathscr{L}_{\mathbf{H}} \Xi^{0}_{(L)a'}) H^{\mu}_{a'} \frac{\partial H^{\lambda}_{a}}{\partial X^{\mu}_{a'}}.$$

The last term is zero because of Eq. (3.30). Thus, we get

$$\mathscr{L}_{H_{(a')}} V^{\lambda}_{(L)a} = -V^{\mu}_{(L)a'} \frac{\partial H^{\lambda}_{a}}{\partial X^{\mu}_{a'}}. \tag{4.23}$$

For any $\bar{X}$ we consider the following functions of the T_b's:

$$V^{\lambda}_{(L)a}(T_b) \equiv V^{\lambda}_{(L)a}[\Omega^{\mu}_{b}(\bar{X}; T_b)]. \tag{4.24}$$

Then Eq. (4.23) tells us that

$$\frac{\partial V^{\lambda}_{(L)a}}{\partial T_{a'}} = -V^{\mu}_{(L)a'} \frac{\partial H^{\lambda}_{a}}{\partial X^{\mu}_{a'}}, \quad \text{at } X^{\mu}_{b} = \Omega^{\mu}_{b}(\bar{X}; T_b). \tag{4.25}$$

Let a be fixed. When all the $T_{a'}$'s ($a' \neq a$) are such that

$$\bar{X}^{0}_{a'} + T_{a'} = \bar{X}^{0}_{a} + T_{a}, \tag{4.26}$$

then we have $V^{\lambda}_{(L)a} = 0$ because of the hypothesis (4.16) and, therefore, also

$$\frac{d^n}{dT_a^n} V^\lambda_{(L)a} = \left(\sum_b \frac{\partial}{\partial T_b} \right)^n V^\lambda_{(L)a} = 0 \quad (n = 1,2,...). \tag{4.27}$$

This formula allows us to express $\partial^n V^\lambda_{(L)a}/\partial T_a^n$ as a homogeneous function of degree one of partial T-derivatives of order $n-1$ of terms of the form $\partial V^\lambda_{(L)a}/\partial T_{a'}$ $(a' \neq a)$. From this and Eq. (4.25) there follows that, when the $T_{a'}$'s satisfy Eq. (4.26), not only $V^\lambda_{(L)a}$ vanish, but also their T-derivatives of any order vanish. Consequently, we have $V^\lambda_{(L)a} = 0$ identically, which completes the proof.

The interpretation of the vectors $\boldsymbol{\eta}_{(L)}$ is clear. They are the infinitesimal generators of the constant-t transformations $y \to \tilde{y} = \tilde{F}_\Lambda(y)$ defined implicitly by the equations

$$\tilde{y} = F_\Lambda \circ e^{T(y;\Lambda)\mathbf{h}}(y), \tag{4.28}$$

with $T(y;\Lambda) = F^0_\Lambda{}_{,}(\tilde{y}) - t$, so that $\tilde{y}^0 = y^0 \equiv t$, i.e., $\tilde{F}_\Lambda$ is the constant-t transformation such that the image of any dynamically possible N-tuple of geometrical trajectories under $\tilde{F}_\Lambda$ is the same as the image under F_Λ (see Fig.2). In general, these transformations $\tilde{F}_\Lambda$ do not constitute a group, but it is clear that when the equations of motion are invariant under G_r then they do constitute one, giving a realization $\Lambda \mapsto \tilde{F}_\Lambda$ of G_r on $\mathbb{R}^{2DN+1}$. A necessary set of conditions for $\Lambda \mapsto \tilde{F}_\Lambda$ to be a realization of G_r is

$$[\boldsymbol{\eta}_{(L)}, \boldsymbol{\eta}_{(M)}] = C^N_{LM} \boldsymbol{\eta}_{(N)}, \tag{4.29}$$

and thus Eqs. (4.29) are *necessary* conditions for the world-line invariance of the equations of motion under G_r. We now amplify this result.

Theorem 2: Let the one-parameter group of time translations, with generator $\xi_{(0)} = \partial/\partial t$, be a subgroup of Eq. (4.1). Then

$$\frac{\partial h^I_a}{\partial t} = 0 \tag{4.30}$$

together with any of the following:

$$\text{(i)} \quad \{[\mathbf{h}, \boldsymbol{\eta}_{(L)}]\}_{t=t_0} = 0, \tag{4.31}$$

$$\text{(ii)} \quad [\boldsymbol{\eta}_{(0)}, \boldsymbol{\eta}_{(M)}] = C^N_{0M} \boldsymbol{\eta}_{(N)}, \tag{4.32}$$

or

$$\text{(iii)} \quad \{[\boldsymbol{\eta}_{(L)}, \boldsymbol{\eta}_{(M)}]\}_{t=t_0} = C^N_{LM} \{\boldsymbol{\eta}_{(N)}\}_{t=t_0} \tag{4.33}$$

constitute a set of necessary and sufficient conditions for the invariance of the equations of motion (3.8) under G_r. (t_0 is arbitrary but fixed; for example, $t_0 = 0$.)

Proof: (1) Equation (4.16) $\Rightarrow$ Eq. (4.30): From $\xi_{(0)} = \partial/\partial t$ we have

$$\boldsymbol{\eta}_{(0)} = -\sum_a h^I_a \frac{\partial}{\partial X^I_a}, \tag{4.34}$$

and thus

$$\mathbf{h} = \frac{\partial}{\partial t} - \boldsymbol{\eta}_{(0)}. \tag{4.35}$$

Consequently,

$$[\mathbf{h}, \boldsymbol{\eta}_{(L)}] = \dot{\boldsymbol{\eta}}_{(L)} - [\boldsymbol{\eta}_{(0)}, \boldsymbol{\eta}_{(L)}], \tag{4.36}$$

where the dot indicates that $\dot{\boldsymbol{\eta}}_{(L)}$ is the vector field whose natural components are the time derivatives of the natural components of $\boldsymbol{\eta}_{(L)}$:

$$\dot{\boldsymbol{\eta}}_{(L)} \equiv \sum_a \frac{\partial \eta^I_{(L)a}}{\partial t} \frac{\partial}{\partial X^I_a}.$$

From Eq. (4.36) there follows that Eq. (4.16) implies

$$\dot{\boldsymbol{\eta}}_{(L)} = [\boldsymbol{\eta}_{(0)}, \boldsymbol{\eta}_{(L)}], \tag{4.37}$$

and taking $L = 0$, we get $\dot{\boldsymbol{\eta}}_{(0)} = 0$, which is equivalent to Eq. (4.30) because of Eq. (4.34).

(2) Equation (4.16) $\Rightarrow$ Eqs. (4.31), (4.32), and (4.33): This is trivial, taking into account that Eq. (4.16) $\Rightarrow$ Eq. (4.29).

(3) Equations (4.30) and (4.31) $\Rightarrow$ Eq. (4.16): First we obtain consequences of Eq. (4.30) alone. From $\boldsymbol{\Xi}_{(0)} = \partial/\partial t$ (on $\mathbb{R}^{2D+1}$) we obtain

$$[\boldsymbol{\Xi}_{(0)}, \boldsymbol{\Xi}_{(L)}] = \dot{\boldsymbol{\Xi}}_{(L)} \quad (\text{on } \mathbb{R}^{2D+1}),$$

and since

$$[\boldsymbol{\Xi}_{(0)}, \boldsymbol{\Xi}_{(L)}] = C^N_{0L} \boldsymbol{\Xi}_{(N)} \quad (\text{on } \mathbb{R}^{2D+1}),$$

there follows that

$$\dot{\boldsymbol{\Xi}}_{(L)} = C^N_{0L} \boldsymbol{\Xi}_{(N)} \quad (\text{on } \mathbb{R}^{2D+1}). \tag{4.38}$$

On the other hand, from Eqs. (4.17) and (4.30),

$$\frac{\partial \eta^I_{(L)a}}{\partial t}(t, X^J_b) = \frac{\partial \Xi^I_L}{\partial t}(t, X^J_a) - \frac{\partial \Xi^0_{(L)}}{\partial t}(t, X^J_a) h^I_a(t, X^J_b),$$

and using Eq. (4.38), we get

$$\dot{\boldsymbol{\eta}}_{(L)} = C^N_{0L} \boldsymbol{\eta}_{(N)}. \tag{4.39}$$

Now we define

$$\mathbf{V}_{(L)} \equiv [\mathbf{h}, \boldsymbol{\eta}_{(L)}], \tag{4.40}$$

and using Eq. (4.30) and (4.39) [which is a consequence of Eq. (4.30)], we get

$$\dot{\mathbf{V}}_{(L)} = [\mathbf{h}, \dot{\boldsymbol{\eta}}_{(L)}] = [\mathbf{h}, C^N_{0L} \boldsymbol{\eta}_{(N)}] = C^N_{0L} \mathbf{V}_{(N)}. \tag{4.41}$$

Now we use Eq. (4.31) for the first time. If it is satisfied, all $\mathbf{V}_{(L)}$'s vanish for $t = t_0$, but then because of Eq. (4.41) they must vanish identically. Thus we get $\mathbf{V}_{(L)} = 0$, which is precisely Eq. (4.16).

(4) Equations (4.30) and (4.32) $\Rightarrow$ Eq. (4.16): From Eqs. (4.36) and (4.39),

$$[\mathbf{h}, \boldsymbol{\eta}_{(L)}] = C^N_{0L} \boldsymbol{\eta}_{(N)} - [\boldsymbol{\eta}_{(0)}, \boldsymbol{\eta}_{(L)}],$$

and the right hand side vanishes if Eq. (4.32) is satisfied.

(5) Equations (4.30) and (4.33) $\Rightarrow$ Eq. (4.32): We define

$$\boldsymbol{\zeta}_{(L,M)} \equiv [\boldsymbol{\eta}_{(L)}, \boldsymbol{\eta}_{(M)}] - C^N_{LM} \boldsymbol{\eta}_{(N)}. \tag{4.42}$$

Then, using Eq. (4.39), one easily obtains

$$\dot{\boldsymbol{\zeta}}_{(L,M)} = (C^N_{0L} C^P_{MN} + C^N_{M0} C^P_{LN} + C^N_{LM} C^P_{0N}) \boldsymbol{\eta}_{(P)} + C^N_{0L} \boldsymbol{\zeta}_{(M,N)} - C^N_{0M} \boldsymbol{\zeta}_{(L,N)}.$$

Because of the Jacobi relations between the structure constants[9] we get

$$\dot{\boldsymbol{\zeta}}_{(L,M)} = C^N_{0L} \boldsymbol{\zeta}_{(M,N)} - C^N_{0M} \boldsymbol{\zeta}_{(L,N)}, \tag{4.43}$$

and therefore if the $\boldsymbol{\zeta}_{(L,M)}$'s vanish simultaneously for $t = t_0$, i.e., if Eq. (4.33) is satisfied, they vanish identically and we have Eq. (4.29). Taking $L = 0$, we get Eq. (4.32). This completes the proof of the theorem.

V. CONCLUSIONS

We have obtained necessary and sufficient conditions for the world-line invariance under a group of transformations of space–time of a dynamical system with Newtonian causality. These conditions were expressed first in terms of vectors on $(\mathbb{R}^{2D+1})^N$ [Eq. (4.14)] and then in terms of vectors on $\mathbb{R}^{2D+1}$ [Eq. (4.16)]. When the compact equations (4.16) are expressed explicitly in terms of the acceleration functions μ_a^i and the generators of the realization of the group on space–time $\xi_{(L)}$, they take the form [omitting the subindex (L) of each of the ξ's, since it is common to all of them]

$$\begin{aligned}
&\xi_a^0 \frac{\partial \mu_a^i}{\partial t} \sum_b \left[\xi_b^j + v_b^j(\xi_a^0 - \xi_b^0)\right] \frac{\partial \mu_a^i}{\partial x_b^j} \\
&\quad + \sum_b \left[\xi_{b,0}^j + v_b^k \xi_{b,k}^j - v_b^j \xi_{b,0}^0 - v_b^j v_b^k \xi_{b,k}^0 \right. \\
&\quad \left. + \mu_b^j(\xi_a^0 - \xi_b^0)\right] \frac{\partial \mu_a^i}{\partial v_b^j} - (\xi_{a,j}^i - v_a^i \xi_{a,j}^0)\mu_a^j \\
&\quad + 2(\xi_{a,0}^0 + v_a^j \xi_{a,j}^0)\mu_a^i - \xi_{a,00}^i - 2v_a^j \xi_{a,0j}^i + v_a^i \xi_{a,00}^0 \\
&\quad + 2v_a^i v_a^j \xi_{a,0j}^0 - v_a^j v_a^k \xi_{a,jk}^i + v_a^i v_a^j v_a^k \xi_{a,jk}^0 = 0 . \qquad (5.1)
\end{aligned}$$

This is a generally nonlinear system of first-order partial differential equations for the functions μ_a^i. Since we have one such set for each of the r values of L, there are DNr equations for the DN unknowns μ_a^i. We have not studied under which conditions this system has solutions where each particle is influenced by all others, nor does this seem to be an easy problem in the general case.[12]

The nonlinear terms are $(\xi_a^0 - \xi_b^0)\mu_b^j(\partial \mu_a^i / \partial v_b^j)$ and therefore the system is linear if and only if $\xi^0(x_a^i, t) = \xi^0(x_b^i, t)$, i.e., iff each $\xi_{(L)}^0$ is at most a function of t, which is equivalent to the condition that the transformations $t \mapsto t' = f^0(t; \Lambda^L)$ should be independent of the spatial coordinates. This is, of course, the case for inhomogeneous Galilei transformations (where $t \mapsto t' = t + A^0$) but not for Poincaré transformations, for example. Furthermore, whether the system is linear or not, it admits the solution $\mu_a^i = 0$ (all particles moving along straight lines with constant velocity; cf. Newton's first law) only if the last six terms on the left hand side vanish. Assuming that the transformations constitute a group, a necessary and sufficient condition for this is that ξ^α's be of the form

$$\xi^\alpha(x) = A^\alpha + B_\beta^\alpha x^\beta + C_\beta x^\alpha x^\beta ,$$

where all the coefficients are constants. In terms of the transformations themselves, this requires that the f^α's be ratios of linear functions of the x^β's all with the same denominator[13]. A particular case is when the transformations are linear, as with the Galilei and Poincaré groups.

In most fundamental physical theories time is assumed to be homogeneous. Therefore, the consideration of dynamical systems which are invariant under time translations is physically relevant, as well as mathematically convenient. For the case of invariance under a group of transformations including time translations as a subgroup we have proved that the conditions (4.16) are equivalent to the commutation relations $[\eta_{(L)}, \eta_{(M)}] = C_{LM}^N \eta_{(N)}$ together with the requirement that the acceleration functions μ_a^i should not depend explicitly on t [i.e., that the dynamical system (3.7) should be autonomous]. Actually, these conditions may be relaxed somewhat (Theorem 2 of Sec. IV.D). In particular, we only need to require the commutation relations for the η's for one particular value of t, for example $t = 0$. This form of the necessary and sufficient conditions for the invariance of the dynamical system as commutation relations between infinitesimal generators for the transformations of initial conditions at $t = 0$ was obtained by Bel[6,2] for the Poincaré group (which, of course, includes time translations), and here this result has been generalized.[14]

In a paper in preparation the conditions for the transformations to be canonical will be studied.

ACKNOWLEDGMENTS

It is a pleasure to thank my Ph.D. thesis adviser, Dr. Peter Havas, for his help and encouragement during the course of this work. My thanks are also due to Dr. Luis Bel, specially for his stimulating monographic course at the Universidad Autónoma de Madrid, on which this paper is partially based. The financial support of Temple University and a Fulbright travel grant are also gratefully acknowledged.

[1] P. Havas and J. Plebański, Bull. Am. Phys. Soc. **5**, 433 (1960).

[2] L. Bel, Ann. Inst. H. Poincaré **14**, 189 (1971); the expression "finitely predictive", which is more general than "predictive" since it only says that the number of initial data which determine the dynamically possible trajectories is finite, was used by Ph. Droz-Vincent, Phys. Scr. **2**, 129 (1970).

[3] Two reviews, with numerous references, are those of P. Havas, in *Statistical Mechanics of Equilibrium and Non-Equilibrium*, edited by J. Meixner (North-Holland, Amsterdam, 1965); and of D.G. Currie and T.F. Jordan, in *Lectures in Theoretical Physics*, edited by W.E. Brittin and A.O. Barut (Gordon and Breach, New York, 1968), Vol. X-A, p. 91.

[4] D.G. Currie, Phys. Rev. **142**, 817 (1966).

[5] R.N. Hill, J. Math. Phys. **8**, 201 (1967).

[6] L. Bel, Ann. Inst. H. Poincaré **12**, 307 (1970).

[7] The treatment in this paper is only local, even if this is not stated explicitly. All functions considered are assumed to be analytic.

[8] The essential points of this subsection were obtained by Ph. Droz-Vincent, Lett. Nuovo Cimento **1**, 839 (1969), and Ref. 2. In particular, he derived the integrability conditions (3.29). However, he considered only relativistic invariance and consequently he parametrized the trajectories by the proper times rather than the coordinate time differences (3.17). Some years before, P. Havas and J. Plebański (unpublished, 1960) had followed an approach similar to the one of this paper, but they stopped short of obtaining Eq. (3.29).

[9] See, for example, L.P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961), Chap. I.

[10] This is essential. If the action of the group is given on $\mathbb{R}^{2DN+1}$ the derivation of the necessary and sufficient conditions (4.16) is trivial. This is what is done in R.N. Hill and E.H. Kerner, Phys. Rev. Lett. **17**, 1156 (1966); and in R.N. Hill, J. Math. Phys. **8**, 1756 (1966). The equations in those papers equivalent to our Eq. (4.16) are Eqs. (2) and (17), respectively.

[11] A proof, which is an adaptation of the proof of Theorem [29.1] in Eisenhart (Ref. 9), is given in A. Salas, "Systems of classical interacting point particles with Newtonian canonicity: world-line invariance and canonicity," Temple University Ph.D. thesis (1978), Appendix 3.

[12]For the case of Poincaré invariance and $N = 2$, it has been proved geometrically that there are such solutions: R. Arens, Arch. Rat. Mech. Anal. **47**, 255 (1972); H.P. Künzle, Symp. Math. **14**, 53 (1974) [cf. the well-known no-interaction theorems (see Ref. 3), which in addition to world-line invariance require the transformations of the Poincaré group to be canonical, and the physical position coordinates x_a^i to be canonical].

[13]V. Fock, *The Theory of Space Time and Gravitation* (Pergamon, New York, 1964), Appendix A.

[14]At first sight it might appear that several results of this paper, notably Theorem 1, are a direct consequence of Lie's fundamental theorems of the theory of continuous groups of transformations, and hence their laborious proof of sufficiency was unnecessary. (For example, the sufficiency of Currie–Hill's conditions for Poincaré invariance was assumed even before it was proved rigorously in Ref. 6.) However, this is not the case, as discussed at the end of Ref. 6.

Time ordered operator cumulants: Statistical independence and noncommutativity

Ronald Forrest Fox
School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 24 January 1979; accepted for publication 27 July 1979)

A cumulant identity for noncommuting operator random processes is discussed. Its validity is negated by exhibiting a counterexample in which statistical independence and noncommutativity are clearly separated.

I. INTRODUCTION

In a variety of contexts, one encounters random processes which are linear combinations of statistically independent random processes. For example, let the process $\tilde{x}$ be given by

$$\tilde{x} = \tilde{y} + \tilde{z} \tag{1}$$

in which $\tilde{y}$ and $\tilde{z}$ are statistically independent. The cumulant functions associated with each of these random processes are defined by

$$\begin{aligned} \exp[K_x(\xi)] &= \langle \exp(\xi\tilde{x})\rangle, \\ \exp[K_y(\xi)] &= \langle \exp(\xi\tilde{y})\rangle, \\ \exp[K_z(\xi)] &= \langle \exp(\xi\tilde{z})\rangle. \end{aligned} \tag{2}$$

Kubo [1] has shown that from (1), and the assumed statistical independence of $\tilde{y}$ and $\tilde{z}$, that the identity

$$\langle \exp(\xi\tilde{x})\rangle = \langle \exp(\xi\tilde{y})\rangle \langle \exp(\xi\tilde{z})\rangle \tag{3}$$

holds. From (3) it follows that the cumulant functions are related by

$$K_x(\xi) = K_y(\xi) + K_z(\xi). \tag{4}$$

In physics, [1] one often encounters random processes which are also operators. Noncommutativity becomes important and ordered exponentials are required. For example, let the operator process $\widetilde{X}$ be given by

$$\widetilde{X} = \widetilde{Y} + \widetilde{Z} \tag{5}$$

in which $\widetilde{Y}$ and $\widetilde{Z}$ are statistically independent operator random processes. If $\exp_Q$ is used to denote an ordered exponential, ordered by the ordering operation Q, then Kubo [1] has proposed the ordered generalization of (3)

$$\langle \exp_Q(\xi\widetilde{X})\rangle = Q(\langle \exp_Q(\xi\widetilde{Y})\rangle \langle \exp_Q(\xi\widetilde{Z})\rangle). \tag{6}$$

In a subsequent critique of the generalized cumulant expansion method for operator processes, Fox [2] asserted that if Q ordering is time ordering, then identity (6) is no longer valid, and he said, "it is not difficult to show that for $\underleftarrow{T}$ ordering (time ordering), counterexamples to Eq. (6) can be constructed." Several valid results derived by Kubo using (6) were rederived by Fox using other methods.

Recently, Apresian [3] has published a paper in which he reviews the generalized cumulant expansion method and in which he rederives one of Fox's results by invoking an identity equivalent with (6) for the case in which Q ordering is time ordering. Apparently, Fox's assertion that counter examples may be *easily constructed* in this case has been insufficiently convincing.

It is the purpose of this paper to exhibit a simple counterexample explicitly. This example makes clear the fact that for operator random processes statistical independence and noncommutativity must both be taken into account during cumulant analysis. Specifically, the counterexample involves an operator random process which is the sum of two *statistically independent* but *noncommutative* operator random processes.

II. TIME ORDERED EXPONENTIALS

Consider the stochastic differential equation [2]

$$\frac{d}{dt}\mathbf{b}(t) = \widetilde{B}(t)\mathbf{b}(t) \tag{7}$$

in which $\widetilde{B}(t)$ is a time-dependent operator random process and $\mathbf{b}(t)$ is a vector. The formal solution to (7) is

$$\mathbf{b}(t) = \underleftarrow{T}\exp\left[\int_0^t ds\widetilde{B}(s)\right]\mathbf{b}(0) \tag{8}$$

in which $\underleftarrow{T}\exp$ denotes the time-ordered exponential which may be defined by

$$\begin{aligned} &\underleftarrow{T}\exp\left[\int_0^t ds\,\widetilde{B}(s)\right] \\ &\quad = 1 + \int_0^t ds\widetilde{B}(s) + \int_0^t ds_1 \int_0^{s_1} ds_2 \widetilde{B}(s_1)\widetilde{B}(s_2) + \cdots \\ &\qquad + \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \\ &\qquad \times \int_0^{s_{n-1}} ds_n\, \widetilde{B}(s_1)\widetilde{B}(s_2)\cdots\widetilde{B}(s_n) + \cdots. \end{aligned} \tag{9}$$

Alternatively, $\underleftarrow{T}$ may be defined by its action on an arbitrary product of integrated operators

$$\begin{aligned} &\underleftarrow{T}\left\{\prod_{j=1}^{k}\int_0^t ds\, M^{(j)}(s)\right\} \\ &\quad = \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{k-1}} ds_k \sum_{\rho} M^{(\rho(1))}(s_1) \\ &\qquad \times M^{(\rho(2))}(s_2)\cdots M^{(\rho(k))}(s_k) \end{aligned} \tag{10}$$

in which Σ_ρ is the sum over all $k!$ permutations of the index set $j = 1,2,\ldots,k$. By expanding the exponential in (8) in the ordinary way, and then applying $\underleftarrow{T}$ to the summands according to (10), Eq. (9) is obtained.

Time-ordered cumulants associated with the operators random process $\widetilde{B}$ are defined by [2]

$$\underleftarrow{T}\exp\left[\int_0^t ds\widetilde{B}(s)\right] \equiv \underleftarrow{T}\exp\left[\sum_{n=1}^{\infty}\int_0^t ds\, G^{(n)}(s)\right]. \tag{11}$$

 0022-2488/79/122467-04$01.00

One may also define the moments of $\tilde{B}$ very naturally by

$$\underset{\leftarrow}{T}\exp\left[\int_0^t ds\tilde{B}(s)\right] = \sum_{m=0}^{\infty}\frac{1}{m!}\left\langle \underset{\leftarrow}{T}\left(\int_0^t ds\tilde{B}(s)\right)^m\right\rangle \equiv 1+\sum_{m=1}^{\infty}\int_0^t ds\, A^{(m)}(s). \tag{12}$$

The moments and the ordered cumulants are know [2] to be related by

$$\int_0^t ds\, A^{(m)}(s) = \sum_{\substack{\text{partitions}\\ \text{of } m}} \underset{\leftarrow}{T}\left\{\prod_{l=1}^{\infty}\frac{1}{m_l!}\left(\int_0^t ds G^{(l)}(s)\right)^{m_l}\right\} \tag{13}$$

in which the "sum over partitions of m" is specified by

$$\sum_{l=1}^{\infty} l m_l = m. \tag{14}$$

This relation may be inverted to obtain the ordered cumulants in terms of sums of products of the moments, and it is known [2,4,5] that for noncommuting operator processes that the inverted relation must be written out with great care.

In particular, if $\langle\tilde{B}(s)\rangle = 0$, then

$$\begin{aligned}
\int_0^t ds\, G^{(1)}(s) &= \int_0^t ds\, A^{(1)}(s) = 0,\\
\int_0^t ds\, G^{(2)}(s) &= \int_0^t ds\, A^{(2)}(s) = \int_0^t ds_1\int_0^{s_2} ds_2\,\langle\tilde{B}(s_1)\tilde{B}(s_2)\rangle,\\
\int_0^t ds\, G^{(3)}(s) &= \int_0^t ds\, A^{(3)}(s) = \int_0^t ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\\
&\quad\times\langle\tilde{B}(s_1)\tilde{B}(s_2)\tilde{B}(s_3)\rangle,\\
\int_0^t ds\, G^{(4)}(s) &= \int_0^t ds\, A^{(4)}(s) - \tfrac{1}{2}\underset{\leftarrow}{T}\left[\left(\int_0^t A^{(2)}(s)\right)^2\right]\\
&= \int_0^t ds_1\int_0^{s_1} ds_2\int_0^{s_2} ds_3\int_0^{s_3} ds_4\{\langle\tilde{B}(s_1)\tilde{B}(s_2)\\
&\quad\times\tilde{B}(s_3)\tilde{B}(s_4)\rangle - \langle\tilde{B}(s_1)\tilde{B}(s_2)\rangle\\
&\quad\times\langle\tilde{B}(s_3)\tilde{B}(s_4)\rangle\\
&\quad - \langle\tilde{B}(s_1)\tilde{B}(s_3)\rangle\,\langle\tilde{B}(s_2)\tilde{B}(s_4)\rangle\\
&\quad - \langle\tilde{B}(s_1)\tilde{B}(s_4)\rangle\langle\tilde{B}(s_2)\tilde{B}(s_3)\rangle\}.
\end{aligned} \tag{15}$$

The assumption that $\langle\tilde{B}(s)\rangle = 0$ has rendered the expressions for the second and third cumulants especially simple, and the expression for the fourth cumulant, while more complicated, is also much more simple that it would be otherwise. The higer-order cumulants become complex very rapidly. Nevertheless, closed form combinatorial expressions are available [2,4,5] although for present purposes they will prove unecessary.

III. A COUNTEREXAMPLE TO (6)

Consider the stochastic differential equation

$$\frac{d}{dt}\Omega(t) = i(\tilde{f}(t)b + \tilde{g}(t)b^{+})\Omega(t) \tag{16}$$

in which $\tilde{f}(t)$ and $\tilde{g}(t)$ are Gaussian random processes of zero mean with non-Markovian second moments and are statistically independent:

$$\begin{aligned}
&\langle\tilde{f}(t)\rangle = 0 = \langle\tilde{g}(t)\rangle,\\
&\langle\tilde{f}(t)\tilde{f}(s)\rangle = F(t-s), \quad \langle\tilde{g}(t)\tilde{g}(s)\rangle = G(t-s),\\
&\langle\tilde{f}(t)\tilde{g}(s)\rangle = 0,
\end{aligned} \tag{17}$$

and b and b^{+} are boson annihilation and creation operators, respectively, satisfying the commutation rule

$$[b,b^{+}] = 1. \tag{18}$$

The statistical independence and the noncommutativity of the two portions of the operator random process in (16) have been explicitly separated.

The averaged solution to (16) may be written

$$\langle\Omega(t)\rangle = \left\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds(f(s)b + \tilde{g}(s)b^{+})\right]\right\rangle\Omega(0). \tag{19}$$

On the right-hand side is the desired quantity and it is tempting to compute it using the Kubo–Apresian identity which in this case reads

$$\begin{aligned}
&\underset{\leftarrow}{T}\exp\left[i\int_0^t ds(\tilde{f}(s)b + \tilde{g}(s)b^{+})\right]\\
&= \underset{\leftarrow}{T}\left(\left\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds\tilde{f}(s)b\right]\underset{\leftarrow}{T}\left(\exp\left[i\int_0^t ds\tilde{g}(s)b^{+}\right]\right)\right\rangle\right).
\end{aligned} \tag{20}$$

It will be shown below that this "identity" is incorrect!

If (20) were correct, it would quickly yield a result because

$$\begin{aligned}
\left\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds\tilde{f}(s)b\right]\right\rangle &= \left\langle \exp\left[i\int_0^t ds\tilde{f}(s)b\right]\right\rangle\\
&= \exp\left[-\int_0^t ds_1\int_0^{s_1} ds_2\, F(s_1-s_2)bb\right],
\end{aligned}$$

and

$$\begin{aligned}
\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds\,\tilde{g}(s)b^{+}\right]\rangle &= \langle\exp\left[i\int_0^t ds\,\tilde{g}(s)b^{+}\right]\rangle\\
&= \exp\left[-\int_0^t ds_1\int_0^{s_1} ds_2\, G(s_1-s_2)b^{+}b^{+}\right].
\end{aligned} \tag{21}$$

In each case, the first step follows because the exponential arguments are commutative with themselves for all times so that $\underset{\leftarrow}{T}$ ordering the exponential isn't required. The second step follows because in the commutative case the second cumulant is exact for a Gaussian process of zero mean, [2] and the exponential argument in each case is merely the second cumulant expression which follows from (15) and (17).

Define $\bar{F}(s_1)$ and $\bar{G}(s_1)$ by

$$\bar{F}(s_1) \equiv \int_0^{s_1} ds_2\, F(s_1-s_2)$$

and

$$\bar{G}(s_1) \equiv \int_0^{s_1} ds_2\, G(s_1-s_2). \tag{22}$$

Therefore, the validity of (20) reduces to the validity of

$$\begin{aligned}
&\left\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds(\tilde{f}(s)b + \tilde{g}(s)b^{+})\right]\right\rangle\\
&= \underset{\leftarrow}{T}\left(\exp\left[-\int_0^t ds\bar{F}(s)bb\right]\exp\left[-\int_0^t ds\,\bar{G}(s)b^{+}b^{+}\right]\right).
\end{aligned} \tag{23}$$

The left-hand side may evaluated using the ordered cumulant expansion:

$$\left\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds(\tilde{f}(s)b + \tilde{g}(s)b^{+})\right]\right\rangle$$

$$= \underset{\leftarrow}{T}\exp\left[\sum_{m=1}^{\infty}\int_0^t ds\,(G^{(m)}(s)\right]. \tag{24}$$

From (15) and (17) it follows that

$$\int_0^t ds\,G^{(1)}(s) = 0,$$

$$\begin{aligned}\int_0^t ds\,G^{(2)}(s) &= -\int_0^t ds_1 \int_0^{s_1} ds_2\,[\,F(s_1-s_2)bb \\ &\quad + G(s_1-s_2)b^{+}b^{+}\,] \\ &= -\int_0^t ds_1\,[\,\bar{F}(s_1)bb + \bar{G}(s_1)b^{+}b^{+}\,].\end{aligned} \tag{25}$$

The Gaussian character[6,7] of $\tilde{f}(s)$ and $\tilde{g}(s)$ and (17) imply that

$$\int_0^t ds\,G^{(3)}(s) = 0. \tag{26}$$

In the Appendix, it is shown that the Gaussian character and (15) also imply

$$\begin{aligned}\int_0^t ds\,G^{(4)}(s) &= \int_0^t ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 \int_0^{s_3} ds_4 \\ &\quad \times \{[G(s_1-s_3)\,F(s_2-s_4) \\ &\quad + 2G(s_1-s_4)\,F(s_2-s_3)]b^{+}b\} \\ &\quad - [F(s_1-s_3)\,F(s_2-s_4) \\ &\quad + F(s_1-s_4)\,G(s_2-s_3)]bb^{+}\}.\end{aligned} \tag{27}$$

These first four cumulants are sufficient to demonstrate that (23) can not be true.

Expanding the right-hand side of (23) up to terms of second order in $\bar{F}$ and $\bar{G}$ yields

$$\begin{aligned}&\underset{\leftarrow}{T}\left(\exp\left[-\int_0^t ds\,\bar{F}(s)bb\right]\exp\left[-\int_0^t ds\,\bar{G}(s)b^{+}b^{+}\right]\right) \\ &= 1 - \int_0^t ds(\,\bar{F}(s)bb + \bar{G}(s)b^{+}b^{+}) + \int_0^t ds_1 \int_0^{s_1} ds_2\,\bar{F}(s_1) \\ &\quad \times \bar{F}(s_2)bbbb + \int_0^t ds_1 \int_0^{s_1} ds_2\,\bar{G}(s_1)\bar{G}(s_2)b^{+}b^{+}b^{+}b^{+} \\ &\quad + \underset{\leftarrow}{T}\left(\left(\int_0^t ds\,\bar{F}(s)bb\right)\left(\int_0^t ds\bar{G}(s)b^{+}b^{+}\right)\right) + \cdots.\end{aligned} \tag{28}$$

The last explicit term in (28) may be rendered as

$$\begin{aligned}&\underset{\leftarrow}{T}\left(\exp\left[\int_0^t ds\,\bar{F}(s)bb\right]\left[\int_0^t ds\,\bar{G}(s)b^{+}b^{+}\right]\right) \\ &= \int_0^t ds_1 \int_0^{s_1} ds_2\,[\,\bar{F}(s_1)\bar{G}(s_2)bbb^{+}b^{+} \\ &\quad + \bar{G}(s_1)\bar{F}(s_2)b^{+}b^{+}bb\,]\end{aligned} \tag{29}$$

according to (10). Now, expanding the left-hand side of (23) up to the same order in $\bar{F}$ and $\bar{G}$ yields instead

$$\begin{aligned}&\langle \underset{\leftarrow}{T}\exp\left[i\int_0^t ds(\tilde{f}(s)b + \tilde{g}(s)b^{+})\right]\rangle \\ &= \underset{\leftarrow}{T}\exp\left[\sum_{m=1}^{\infty}\int_0^t ds G^{(m)}(s)\right] \\ &= 1 + \int_0^t ds\,G^{(2)}(s) + \int_0^t ds G^{(4)}(s) + \int_0^t ds_1 \int_0^{s_1} ds_2 \\ &\quad \times G^{(2)}(s_1)G^{(2)}(s_2) + \cdots \\ &= 1 - \int_0^t ds[\,\bar{F}(s)bb + \bar{G}(s)b^{+}b^{+}\,] + \int_0^t ds G^{(4)}(s) \\ &\quad + \int_0^t ds_1 \int_0^{s_1} ds_2\,[\,\bar{F}(s_1)\bar{F}(s_2)bbbb \\ &\quad + \bar{G}(s_1)\bar{G}(s_2)b^{+}b^{+}b^{+}b^{+} \\ &\quad + \bar{F}(s_1)\bar{G}(s_2)bbbb^{+}b^{+} \\ &\quad + \bar{G}(s_1)\bar{F}(s_2)b^{+}b^{+}bb\,] + \cdots.\end{aligned} \tag{30}$$

Equation (30) contains all the terms of Eq. (29) plus the additional, nonvanishing fourth cumulant contribution which is given explicitly in (27).

The basis for this discrepancy, and consequently for the invalidity of the Kubo–Apresian identity (6), lies in the fact that although the commutator of $\tilde{f}(s)b$ and $\tilde{g}(t)b^{+}$,

$$[\,\tilde{f}(s)b,\,\tilde{g}(t)b^{+}\,] = \tilde{f}(s)\tilde{g}(t), \tag{31}$$

vanishes on the average:

$$\langle \tilde{f}(s)\tilde{g}(t)\rangle = 0, \tag{32}$$

its higher moments do not vanish:

$$\langle \tilde{f}(s)\tilde{g}(t)\tilde{f}(s')\tilde{g}(t')\rangle = F(s-s')G(t-t'). \tag{33}$$

Moreover, the average of products of the commutator with $\tilde{f}$ b and $\tilde{g}\,b^{+}$ do not vanish:

$$\langle \tilde{f}(t)b\tilde{f}(t')\tilde{g}(s')\tilde{g}(s)b^{+}\rangle = F(t-t')G(s'-s)bb^{+}. \tag{34}$$

It is precisely such relations which give rise to a nonvanishing fourth cumulant $G^{(4)}$ in the Appendix.

IV. CONCLUSION

In the case of operator random processes which do not commute, statistical independence alone will not eliminate higher-order correlation corrections which have their origin in the noncommutativity. Consequently, the Kubo–Apresian identity is not universally true and can not be invoked in proofs of other results. That such considerations are not purely academic is illustrated by the occurrence of equations with precisely the structure of (16) in the study of the coupling of atomic electrons with phonons.[6]

ACKNOWLEDGMENTS

This work was supported by NSF grant PHY77-07372. The author is indebted to A. Maltsev for bringing the work of L. Apresian to his attention.

APPENDIX: DERIVATION OF (27)

The expression for $\int_0^t ds\,G^{(4)}(s)$ given in (15) may be used if

$$\tilde{B}(s) \equiv \tilde{f}(s)b + \tilde{g}(s)b^{+} \tag{A1}$$

is substituted. A self-explanatory short-hand rendering of this expression is

$$\begin{aligned}\int_0^t ds\,G^{(4)}(s) &= \int_0^t ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 \int_0^{s_3} ds_4\,\{\langle 1\ 2\ 3\ 4\rangle \\ &\quad - \langle 1\ 2\rangle\langle 3\ 4\rangle - \langle 1\ 3\rangle\langle 2\ 4\rangle \\ &\quad - \langle 1\ 4\rangle\langle 2\ 3\rangle\}.\end{aligned} \tag{A2}$$

The Gaussian property implies[7] that $\langle 1\ 2\ 3\ 4\rangle$ may be re-

duced to products of pair correlations, but only if noncommutativity is respected:

$$\langle 1\ 2\ 3\ 4\rangle = \langle 1\ 2\rangle\langle 3\ 4\rangle + \langle 1\ \overbrace{\langle 2\ 3\rangle}\ 4\rangle + \langle 1\ \underbrace{\langle 2\ 3\rangle\ 4}\rangle, \tag{A3}$$

where the bars in the last two terms of the right-hand side indicate which correlation pairs are involved. Therefore, (A2) reduces to

$$\begin{aligned}\int_0^t ds G^{(4)}(s) = \int_0^t ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 \int_0^{s_3} ds_4 \\ \times\{\langle 1\ \overbrace{\langle 2\ 3\rangle}\ 4\rangle - \langle 1\ 3\rangle\langle 2\ 4\rangle \\ + \langle 1\ \underbrace{\langle 2\ 3\rangle\ 4}\rangle - \langle 1\ 4\rangle\langle 2\ 3\rangle\}.\end{aligned} \tag{A4}$$

Using (A1) gives

$$\begin{aligned}\langle 1\ 3\rangle\langle 2\ 4\rangle &= [F(s_1-s_3)bb + G(s_1-s_3)b^+b^+] \\ &\quad \times [F(s_2-s_4)bb + G(s_2-s_4)b^+b^+], \\ \langle 1\ 4\rangle\langle 2\ 3\rangle &= [F(s_1-s_4)bb + G(s_1-s_4)b^+b^+] \\ &\quad \times [F(s_2-s_3)bb + G(s_2-s_3)b^+b^+], \\ \langle 1\ \overbrace{\langle 2\ 3\rangle}\ 4\rangle &= F(s_1-s_3)[F(s_2-s_4)bbbb \\ &\quad + G(s_2-s_4)bb^+bb^+] \\ &\quad + G(s_1-s_3)[F(s_2-s_4)b^+bb^+b \\ &\quad + G(s_2-s_4)b^+b^+b^+b^+], \\ \langle 1\ \underbrace{\langle 2\ 3\rangle\ 4}\rangle &= F(s_1-s_4)[F(s_2-s_3)bbbb \\ &\quad + G(s_2-s_3)bb^+b^+b] \\ &\quad + G(s_1-s_4)[F(s_2-s_3)b^+bbb^+ \\ &\quad + G(s_2-s_3)b^+b^+b^+b^+].\end{aligned} \tag{A5}$$

From (18) it follows that

$$\begin{aligned}bb^+bb^+ &= bbb^+b^+ - bb^+, \\ b^+bb^+b &= b^+b^+bb + b^+b, \\ bb^+b^+b &= bbb^+b^+ - 2bb^+, \\ b^+bbb &= b^+b^+bb + 2b^+b.\end{aligned} \tag{A6}$$

Using (A.6) in (A5) yields the identities

$$\begin{aligned}\langle 1\ \underbrace{\langle 2\ 3\rangle\ 4}\rangle &= \langle 1\ 3\rangle\langle 2\ 4\rangle - F(s_1-s_3)G(s_2-s_4)bb^+ \\ &\quad + G(s_1-s_3)F(s_2-s_4)b^+b,\end{aligned} \tag{A7}$$

$$\begin{aligned}&\langle 1\ \underbrace{\langle 2\ 3\rangle\ 4}\rangle \\ &= \langle 1\ 4\rangle\langle 2\ 3\rangle - 2F(s_1-s_4)G(s_2-s_3)bb^+ \\ &\quad + 2G(s_1-s_4)F(s_2-s_3)b^+b.\end{aligned}$$

Insertion of (A7) into (A4) results in (27).

[1] R.Kubo, J. Phys. Soc. Japan **17**, 1100 (1962).
[2] R.F. Fox, J. Math. Phys. **17**, 1148 (1976).
[3] L.A. Apresian, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **21**, 698 (1978).
[4] N.G. van Kampen, Physica **74**, 215, 239 (1974).
[5] G.W. Ford, "Notes on Stochastic Differential Equations," unpublished (1974).
[6] R.F. Fox, Phys. Rep. C **48**, 179 (1978).
[7] R.F. Fox, J. Math. Phys. **16**, 289 (1975), Sec 3.

Solution of the wave equation for the logarithmic potential with application to particle spectroscopy

H. J. W. Müller-Kirsten and S. K. Bose[a)]
Department of Physics, University of Kaiserslautern, 6750 Kaiserslautern, Germany

(Received 23 October 1978)

We present an almost complete solution of the Schrödinger equation for a logarithmic potential. In particular we obtain two pairs of high-energy asymptotic expansions of the boundstate eigenfunctions together with a corresponding expansion of the eigenvalue determined by the secular equation. We also obtain a pair of uniformly convergent solutions and a pair of uniform asymptotic expansions. Various properties of the solutions and eigenvalues are examined, including the scattering problem of the cut-off potential and the behavior of Regge trajectories. Finally the relevance of these investigations to the spectroscopy of heavy quark composites is discussed. In particular we point out that the relevance of the logarithmic potential can be tested only if more than two consecutive energy levels are known. In a separate paper the methods outlined here are applied to quark-confining potentials of the generalized power type.

1. INTRODUCTION

The states $\Upsilon(9.4\,\mathrm{GeV}/c^2)$ and $\Upsilon'(10.0\,\mathrm{GeV}/c^2)$ recently discovered at Fermilab[1] in the dimuon mass spectrum of the reaction 400 GeV p + nucleus$\rightarrow\mu^+\mu^-$ + anything led to the observation that their mass difference (approximately $0.6\,\mathrm{GeV}/c^2$) is about the same as that between $\psi(3.1\,\mathrm{GeV}/c^2)$ and $\psi'(3.7\,\mathrm{GeV}/c^2)$. Quigg and Rosner[2] therefore raised the question whether the quark–antiquark binding potential is such that this mass difference is independent of the effective mass of the constituent quarks. Using scaling arguments they found that a potential of this type, giving a level spacing which is independent of the quark mass, is the logarithmic potential. Investigating its spectrum Quigg and Rosner found that the logarithmic potential is not ruled out as a candidate for the quark–antiquark interaction provided a $4s$ charmonium level is found near $4.2\,\mathrm{GeV}/c^2$.

The study of quark confining potentials has received new impetus by the discovery of the Υ, Υ' states. In view of forthcoming high energy e^+e^- colliding beam experiments which can be expected to produce new and perhaps unusual data, it is of interest to study the quark–antiquark force on a more general basis and to ask if the data (e.g., level spacings and leptonic decay rates) can be used to infer information on the nature of the quark–antiquark binding potential. With this question in mind Quigg and Rosner[3] and Thacker[4] considered a general power potential and the inverse scattering problem with applications to ψ and Υ families. The investigations of Quigg and Rosner[2,3] are based on scaling arguments and WKB approximations for the wavefunctions. However, for a theoretical exploration it is desirable to have more complete solutions and explicit expressions for energy eigenvalues and Regge trajectories.[5] As Quigg and Rosner[6] have shown, in the case of some potentials, an explicit knowledge of eigenvalues allows a simple evaluation of the S wave bound state wavefunction at the origin and hence the calculation of leptonic decay widths.

In the following we consider in detail the logarithmic potential. We derive various types of solutions of the wavefunctions as well as explicit asymptotic expansions for the energy eigenvalues and Regge trajectories. Our methods of solution are very general and parallel the methods of solution of more complicated standard differential equations such as the Mathieu equation, as a comparison with the relevant literature[7] reveals. In Secs. 2 and 3 we derive two pairs of high energy asymptotic expansions for the discrete eigenfunctions together with the corresponding asymptotic expansion for the eigenvalues. In Secs. 4 and 5 we derive a pair of uniformly convergent expansions for the solutions—this has been obtained previously by Gesztesy and Pittner[8]—and a pair of uniform asymptotic expansions, and we discuss the relevance of these solutions for the scattering problem of the cut-off potential. In Sec. 6 then, we investigate the physical implications of the asymptotic expansions of the energy eigenvalues and Regge trajectories. In particular we demonstrate the near-independence of the level shift of the quark masses and obtain an explicit expression for the bound state wavefunction at the origin. Finally we discuss the relevance of the logarithmic potential for particle spectroscopy.

In a separate publication we repeat the methods developed here for the generalized power potential.

2. A FIRST PAIR OF ASYMPTOTIC EIGENSOLUTIONS

We consider the Schrödinger equation for the unscreened logarithmic potential

$$V(r) = g\ln(r/r_0), \quad g>0.$$

Separating off the motion of the center of mass in the usual way, we obtain the radial wave equation for the relative motion of two particles of masses m_1, m_2,

$$\frac{d^2\psi}{dr^2} + \frac{2\mu}{\hbar^2}\left(E - \frac{l(l+1)\hbar^2}{2\mu r^2} - V(r)\right)\psi = 0, \qquad (1)$$

where, as usual, $\Psi\sim(1/r)\psi(r)$. $P_l^m(\cos\theta)\cdot e^{im\varphi}$.
$\mu = m_1m_2/(m_1+m_2)$ is the reduced mass of the two particles and r is their separation.

[a)] A.V. Humboldt foundation fellow on leave from the University of Gorakhpur.

 0022-2488/79/122471-10$01.00

Inserting the potential, we have the equation

$$\frac{d^2\psi}{dr^2} + \left(\alpha - \beta \ln r - \frac{\gamma}{r^2}\right)\psi = 0, \tag{2}$$

where we have set

$$\alpha = 2\mu E'/\hbar^2, \quad E' = E + g \ln r_0, \tag{3}$$

$$\beta = 2\mu g/\hbar^2, \quad \gamma = l(l+1).$$

Next we set

$$r = e^{z-c} \quad (-\infty < z < \infty), \tag{4}$$

where c is a constant. Setting also

$$\psi = e^{(z-c)/2}\phi \tag{5}$$

and choosing

$$c = -\alpha/\beta, \tag{6}$$

we obtain our basic equation

$$\frac{d^2\phi}{dz^2} + [-L^2 + U(z)]\phi = 0, \tag{7}$$

where

$$L^2 = \gamma + \tfrac{1}{4} = (l + \tfrac{1}{2})^2 \tag{8}$$

and

$$U(z) = -z\beta e^{2\alpha/\beta} e^{2z}. \tag{9}$$

Our next step is to find *that* value of z, say z_0, for which $U(z)$ becomes maximal. In the vicinity of this maximum $U(z) - L^2$ can become positive and the solutions therefore oscillatory as required for the existence of eigenvalues. Simple differentiation yields the value

$$z_0 = -\tfrac{1}{2}.$$

It may be noted that the spectrum of bound states lies in the domain $-\infty < E < \infty$ since $V(r) \to -\infty$ for $r \to 0$ and $V(r) \to +\infty$ for $r \to \infty$.

Expanding $U(z)$ in the neighborhood of the maximum at z_0 we obtain

$$U(z) = U(z_0) + \sum_{i=2}^{\infty} \frac{(z - z_0)^i}{i!} U^{(i)}(z_0), \tag{10}$$

where

$$U^{(i)}(z_0 = -\tfrac{1}{2}) = -\beta e^{(2\alpha-\beta)/\beta}(i-1)2^{i-1} \tag{11}$$

for $i = 0,1,2,\cdots$. For $i = 0$ this expression is positive, for $i = 1$ it is zero, and for $i > 1$ it is negative [as required for a maximum of $U(z)$ at $z = z_0$ for $\beta > 0$]. We now set

$$h^4 = 4\beta e^{(2\alpha-\beta)/\beta} \equiv 4\epsilon/e, \tag{12}$$

i.e.,

$$h^2 = [-2U^{(2)}(z_0)]^{1/2} = \frac{2r_0}{\hbar}(2\mu g)^{1/2} \exp\left(\frac{E}{g} - \frac{1}{2}\right).$$

(It will be convenient at times to use ϵ instead of h as the parameter of our expansions.)

We now change the independent variable in (7) to

$$\omega = h(z - z_0). \tag{13}$$

The equation then becomes

$$\frac{d^2\phi}{d\omega^2} + \left(\frac{1}{h^2}(-L^2 + \tfrac{1}{8}h^4) - \tfrac{1}{4}\omega^2\right)\phi = \sum_{i=3}^{\infty} \frac{(i-1)2^{i-3}}{i!} \cdot \frac{\omega^i}{h^{i-2}}\phi. \tag{14}$$

For large values of h the right-hand side of this equation may—to a first approximation—be neglected. The corresponding behavior of the "eigenvalues" $(1/h^2)(-L^2 + \frac{1}{8}h^4)$ can then be determined by comparing the equation with the equation of parabolic cylinder functions. The solutions are square-integrable only if

$$\frac{1}{h^2}(-L^2 + \tfrac{1}{8}h^4) = \tfrac{1}{2}q,$$

where q is an odd integer, i.e., $2n + 1$, $n = 0,1,2,\cdots$ [provided the wavefunction is required to vanish at infinity; otherwise—e.g., if the potential is cut off at some value $r = R_0 - q$ is only approximately an odd integer q_0, i.e., $q = q_0 + O(1/R_0)$]. For the complete solution we set

$$\frac{1}{h^2}(-L^2 + \tfrac{1}{8}h^4) = \tfrac{1}{2}q + \frac{\Delta}{h}. \tag{15}$$

The quantity Δ in (15) remains to be determined. Substituting (15) in (14) we have an equation which can be written

$$\mathscr{D}_q\phi = \frac{2\Delta}{h}\phi + \sum_{i=3}^{\infty} \frac{(1-i)2^{i-2}}{i!} \cdot \frac{\omega^i}{h^{i-2}}\phi, \tag{16}$$

where

$$\mathscr{D}_q \equiv -2\frac{d^2}{d\omega^2} - q + \tfrac{1}{2}\omega^2. \tag{17}$$

Equation (16) is now in a form suitable for the application of our perturbation method. To a first approximation $\phi = \phi^{(0)}$ is simply a parabolic cylinder function $D_{(q-1)/2}(\omega)$, i.e.,

$$\phi^{(0)} = \phi_q = D_{(q-1)/2}(\omega), \quad \mathscr{D}_q\phi_q = 0, \tag{18}$$

where

$$D_{(q-1)/2}(\omega) = 2^{(q-3)/4} e^{-\omega^2/4} \Phi\left(\frac{3-q}{4}, \frac{3}{2}; \frac{\omega^2}{2}\right),$$

Φ being a confluent hypergeometric function. The function ϕ_q is well known to obey the recurrence formula

$$\omega\phi_q = (q, q+2)\phi_{q+2} + (q, q-2)\phi_{q-2}, \tag{19}$$

where

$$(q, q+2) = 1, \quad (q, q-2) = \tfrac{1}{2}(q-1). \tag{20}$$

For higher powers we have

$$\omega^i\phi_q = \sum_{j=2i}^{-2i} S_i(q, j)\phi_{q+j}. \tag{21}$$

and a recurrence relation can be written down for the coefficients S. The first approximation $\phi = \phi^{(0)}$ then leaves uncompensated terms on the right-hand side of (16) amounting to

$$R_q^{(0)} = \frac{2\Delta}{h}\phi_q + \sum_{i=3}^{\infty} \frac{(1-i)2^{i-2}}{i!h^{i-2}} \omega^i\phi_q$$

$$\equiv \frac{2\Delta}{h}\phi_q - \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=-2i}^{2i}\tilde{S}_i(q,j)\phi_{q+j}, \tag{22}$$

where we have set

$$\tilde{S}_i(q,j)\equiv\frac{(i-1)2^{i-2}}{i!}S_i(q,j). \tag{23}$$

We now rewrite (22) in the form

$$R_q^{(0)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2i}^{-2i}[q,q+j]_i\,\phi_{q+j}(\omega), \tag{24}$$

where

$$[q,q]_3 = 2\Delta - \tilde{S}_3(q,0)$$

and for $j\neq 0$

$$[q,q+j]_3 = -\tilde{S}_3(q,j), \tag{25}$$

and for $i>3$

$$[q,q+j]_i = -\tilde{S}_i(q,j).$$

Since

$$\mathscr{D}_{q+j} = \mathscr{D}_q - j, \quad \mathscr{D}_q\phi_{q+j} = j\phi_{q+j}$$

a term $\mu\phi_{q+j}$ in $R_q^{(0)}$ may be removed by adding to $\phi^{(0)}$ the contribution $\mu\phi_{q+j}/j$ except, of course, when $j=0$. Thus the next-order contribution of ϕ becomes

$$\phi^{(1)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q,q+j]_i}{j}\phi_{q+j}(\omega). \tag{26}$$

In its turn this contribution leaves uncompensated

$$R_q^{(1)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q,q+j]_i}{j}R_{q+j}^{(0)}$$

and yields the next contribution

$$\phi^{(2)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q,q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'-2}} \times \sum_{\substack{j'=2i'\\ j'+j'\neq 0}}^{-2i'}\frac{[q+j,q+j+j']_i}{j+j'}\phi_{q+j+j'}. \tag{27}$$

Proceding in this way we obtain the solution $\phi=\phi^{(0)}+\phi^{(1)}+\phi^{(2)}+\cdots$ which is an asymptotic expansion in descending powers of h, valid for $\frac{1}{2}\omega^2<1$ [convergence of the hypergeometric expansion in $D_{(q-1)/2}(\omega)$]. Thus

$$z - z_0 < 2^{1/2}/h \text{ or } z \text{ around } z_0. \tag{28}$$

Together with this solution we obtain an eigenvalue equation from which Δ in (15) follows. The latter is obtained by setting equal to zero the sum of the terms in ϕ_q in $R_q^{(0)}, R_q^{(1)},\cdots$ which have been unaccounted for so far. Thus

$$0 = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}[q,q]_i + \sum_{i=3}^{\infty}\frac{1}{h^{i-2}} \times \sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q,q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'-2}}[q+j,q]_{i'} + \cdots \tag{29}$$

or

$$0 = \frac{1}{h}[q,q]_3 + \frac{1}{h^2}\left([q,q]_4 + \sum_{\substack{j=6\\ j\neq 0}}^{-6}\frac{[q,q+j]_3}{j}[q+j,q]_3\right) + O\left(\frac{1}{h^3}\right). \tag{30}$$

This is the equation from which Δ and hence the eigenvalues are determined. As in our previous investigation[5] of a power potential, the expansion is much simpler than may appear on first sight because [owing to the fact that (19) does not contain terms in $\phi_{q+1},\phi_q,\phi_{q-1}$] many terms are zero, e.g., $\tilde{S}_3(q,0)$. Thus

$$2h\Delta = [\tilde{S}_4(q,0) - \tfrac{1}{6}\tilde{S}_3(q,6)\tilde{S}_3(q+6,-6) + \tfrac{1}{6}\tilde{S}_3(q,-6)\tilde{S}_3(q-6,6) - \tfrac{1}{2}\tilde{S}_3(q,2)\tilde{S}_3(q+2,-2) + \tfrac{1}{2}\tilde{S}_3(q,-2)\tilde{S}_3(q-2,2)] + O\left(\frac{1}{h}\right)$$

$$= \frac{-1}{2^2 3^2}(33q^2+1) + \frac{q}{2^4 3^2 h^2}(35q^2-1) + O\left(\frac{1}{h^4}\right).$$

Substituting Δ into (15) we obtain

$$(l+\tfrac{1}{2})^2 = \frac{h^4}{8} - \tfrac{1}{2}qh^2 - \frac{1}{2^3 3^2}(33q^2+1) - \frac{q(35q^2-1)}{2^5 3^3 h^2} + O\left(\frac{1}{h^4}\right). \tag{31}$$

[The coefficient of the term of $O(1/h^2)$ is the contribution of S.K. Bose who collaborated in the initial stage of this work.]

We have thus obtained a large-h asymptotic expansion of an eigenfunction of the Schrödinger equation for the logarithmic potential. This expansion is valid in the region around $z=z_0$ or $\ln r = \alpha/\beta - \frac{1}{2}$. A second, linearly independent solution in the same domain is obtained by the interchanges $\omega\to\pm i\omega$, $q\to -q$, $h\to\pm ih$.

3. A SECOND PAIR OF ASYMPTOTIC EIGENSOLUTIONS

Our next objective is to derive a second pair of large-h asymptotic expansions for the eigenfunctions of the logarithmic potential which is valid in regions of large z where the expansions obtained above are no longer applicable [see (28)]. Of course, the corresponding eigenvalue expansion will be identical with (31) above. It is convenient to use the parameter ϵ instead of h, i.e.,

$$\epsilon = \frac{e}{4}h^4.$$

Our starting point is Eq. (7) in which we insert for L^2 the expressson (15) in terms of the quantity $\Delta' = \Delta h$ which is to be determined by iteration. We then have the equation

$$\frac{d^2\phi}{dz^2} - \epsilon\left(\frac{1}{2e} + ze^{2z}\right)\phi + \left(\frac{q}{e^{1/2}}\epsilon^{1/2} - \Delta'\right)\phi = 0. \tag{32}$$

We now set

$$\phi = \chi(z)\exp\left(\pm\epsilon^{1/2}\int^z [v(z)]^{1/2}\,dz\right), \tag{33}$$

where

$$v(z) = \frac{1}{2e} + ze^{2z} \tag{34}$$

and $\chi(z)$ satisfies

$$\frac{d^2\chi}{dz^2} \pm 2\epsilon^{1/2}v^{1/2}(z)\frac{d\chi}{dz} \pm \tfrac{1}{2}\epsilon^{1/2}\frac{v'(z)}{v^{1/2}(z)}\chi + \left(\frac{q}{e^{1/2}}\epsilon^{1/2} - \Delta'\right)\chi = 0. \tag{35}$$

From now on we consider only the equation for the upper signs. The equation for the lower signs leads to another solution which can be obtained from the solution we shall derive by changing the signs of $\epsilon^{1/2}$ and q throughout. Thus, choosing the upper signs in (35), we can rewrite the equation in the form

$$\mathscr{D}_q\chi = \left(\frac{e}{\epsilon}\right)^{1/2}\left(\frac{d^2\chi}{dz^2} - \Delta'\chi\right), \tag{36}$$

where

$$\mathscr{D}_q \equiv -2e^{1/2}v^{1/2}\frac{d}{dz} - \frac{e^{1/2}}{2}\cdot\frac{v'(z)}{v^{1/2}} - q. \tag{37}$$

Since we know that Δ' is at most of $O(0)$ in ϵ for $\epsilon\to\infty$ we can—to a first approximation—neglect the terms on the right-hand side of (36) and write for the solution to that order

$$\chi^{(0)} = \chi_q, \tag{38}$$

where χ_q is the solution of

$$\mathscr{D}_q\chi_q = 0, \tag{39}$$

i.e.,

$$\chi_q(z) = \frac{C}{v^{1/2}}\exp\left(-\frac{q}{2e^{1/2}}\int^z\frac{dz}{v^{1/2}(z)}\right), \tag{40}$$

where C is an overall multiplicative constant which we shall ignore except in the context of normalization.

Proceeding as in the derivation of our first solution, we evaluate $d^2\chi_q/dz^2$ and obtain

$$\frac{d^2\chi_q}{dz^2} - \Delta'\chi_q = \left\{-\Delta' + \frac{5}{16}\frac{v'^2}{v^2} + \frac{qv'}{2e^{1/2}v^{3/2}} + \frac{q^2}{4e}\cdot\frac{1}{v} - \frac{v''}{4v}\right\}\chi_q. \tag{41}$$

Looking at the solution (40) we observe the following relations

$$\frac{\chi_{q+j}}{\chi_q} = \left(\frac{\chi_{q+1}}{\chi_q}\right)^j, \quad \frac{\chi_{q+j}}{\chi_q} = \frac{\chi_q}{\chi_{q-j}}. \tag{42}$$

Further, since

$$\mathscr{D}_{q+j} = \mathscr{D}_q - j$$

and

$$\mathscr{D}_q\chi_{q+j} = j\chi_{q+j} \tag{43}$$

it is desirable to reexpress (41) as a sum over various χ_{q+j} because then the perturbation procedure becomes particularly simple.[9] This type of expansion is simply an expansion in terms of eigenfunctions, such as (e.g.) a Fourier expansion. In order to derive such an expansion it is not possible to use a simple power expansion of the integrand and integral in the relation

$$x \equiv \frac{\chi_{q-1}}{\chi_q} = \exp\left(\frac{1}{2e^{1/2}}\int^z\frac{dz}{v^{1/2}}\right) \tag{44}$$

because in that case reversion of the resulting series leads to divergent sums for each coefficient of the sum over powers of x. The correct step is to use an expansion around $z = z_0 = -\frac{1}{2}$ for which

$$v(z_0) = 0 \quad \textit{and} \quad v'(z_0) = 0.$$

Then

$$v(z) = \sum_{i=2}^{\infty}\frac{(z+\frac{1}{2})^i}{i!}v^{(i)}(z_0), \tag{45}$$

where for $i = 2,3,\cdots$

$$v^{(i)}(z_0 = -\tfrac{1}{2}) = \frac{(i-1)2^{i-1}}{e}. \tag{46}$$

We then have (apart from an additive constant)

$$\frac{1}{2e^{1/2}}\int^z\frac{dz}{v^{1/2}} = \tfrac{1}{2}\ln(z+\tfrac{1}{2}) + \sum_{i=1}^{\infty}\frac{\gamma_i}{2i}(z+\tfrac{1}{2})^i, \tag{47}$$

where

$$\gamma_1 = -\tfrac{2}{3}, \quad \gamma_2 = \tfrac{1}{6}, \quad \gamma_3 = -\tfrac{1}{135},\cdots.$$

Expression (47) can now be substituted into (44) and the resulting expansion in powers of $(z+\frac{1}{2})$ can be reversed. We quote the following two expansions which have to be used repeatedly in subsequent steps:

$$(z+\tfrac{1}{2})^{1/2} = \sum_{i=0}^{\infty}d_{2i+1}\frac{\chi_{q-(2i+1)}}{\chi_q}, \tag{48}$$

where

$$d_1 = 1, \quad d_3 = \tfrac{1}{3}, \quad d_5 = \tfrac{17}{72},\cdots$$

and

$$(z+\tfrac{1}{2})^{-2} = \sum_{i=-2,1,0}^{\infty}f_{2i}\frac{\chi_{q+2i}}{\chi_q},$$

where

$$f_4 = 1, \quad f_2 = -\tfrac{4}{3}, \quad f_0 = -\tfrac{1}{6},\cdots.$$

Inserting (48) into (45) and inverting the series we obtain

$$\frac{1}{v(z)} = e\sum_{i=2,1,0}^{\infty}\delta_{2i}\frac{\chi_{q+2i}}{\chi_q}, \tag{49}$$

where

$$\delta_4 = 1, \quad \delta_2 = -\tfrac{8}{3}, \quad \delta_0 = \tfrac{11}{6}, \cdots.$$

In a similar way we find

$$\frac{v'^2}{v^2} = \sum_{i=2,1,0}^{-\infty} \tau_{2i} \frac{\chi_{q+2i}}{\chi_q}, \tag{50a}$$

where

$$\tau_4 = 4, \quad \tau_2 = 0, \quad \tau_0 = -\tfrac{2}{9}, \cdots,$$

$$\frac{v''}{v} = \sum_{i=2,1,0}^{-\infty} \epsilon_{2i} \frac{\chi_{q+2i}}{\chi_q} \tag{50b}$$

where

$$\epsilon_4 = 2, \quad \epsilon_2 = \tfrac{8}{3}, \quad \epsilon_0 = -\tfrac{1}{3}, \cdots,$$

and

$$\frac{v'}{v^{3/2}} = e^{1/2} \sum_{i=2,1,0}^{-\infty} \kappa_{2i} \frac{\chi_{q+2i}}{\chi_q}, \tag{50c}$$

where

$$\kappa_4 = 2, \quad \kappa_2 = -\tfrac{8}{3}, \quad \kappa_0 = 0, \cdots.$$

These expansions can now be substituted in (41). Then

$$\frac{d^2\chi_q}{dz^2} - \Delta'\chi_q = \sum_{i=2,1,0}^{-\infty} (q,q+2i)\chi_{q+2i}, \tag{51}$$

where for $i \neq 0$:

$$(q,q+2i) = \tfrac{5}{16}\tau_{2i} + \frac{q}{2}\kappa_{2i} + \frac{q^2}{4}\delta_{2i} - \tfrac{1}{4}\epsilon_{2i}$$

and for $i = 0$:

$$(q,q) = -\Delta' + \tfrac{5}{16}\tau_0 + \frac{q}{2}\kappa_0 + \frac{q^2}{4}\delta_0 - \tfrac{1}{4}\epsilon_0$$

$$= -\Delta' + \tfrac{1}{72}(33q^2 + 1). \tag{52}$$

Thus the first approximation $\chi^{(0)} = \chi_q$ leaves uncompensated on the right-hand side of (36) a sum of terms amounting to

$$R_q^{(0)} = \left(\frac{e}{\epsilon}\right)^{1/2} \sum_{i=2,1,0}^{-\infty} (q,q+2j)\chi_{q+2j}. \tag{53}$$

Using (43) we see that these terms can be taken care of by adding to $\chi^{(0)}$ the next order contribution

$$\chi^{(1)} = \left(\frac{e}{\epsilon}\right)^{1/2} \sum_{\substack{j=2,1,\cdots \\ j\neq 0}}^{-\infty} \frac{(q,q+2j)}{2j} \chi_{q+2j} \tag{54}$$

excluding, of course, the term in χ_q. The coefficient of χ_q in (53) set equal to zero, i.e.,

$$(q,q) = 0$$

yields an expression for Δ' (to the same order of approximation) which is identical with the expression obtained previously for our other type of eigenvalue expansion. This result therefore reassures us, that our approach and its solutions are *correct*. The complete solution is obtained in our standard fashion[7,9] leading to the sum

$$\chi = \chi^{(0)} + \chi^{(1)} + \chi^{(2)} + \cdots$$

in powers of $(e/\epsilon)^{1/2}$. The corresponding equation for Δ' and thus the eigenvalues is

$$0 = (q,q) + \left(\frac{e}{\epsilon}\right)^{1/2} \sum_{\substack{j=2,1,\cdots \\ j\neq 0}}^{-\infty} \frac{(q,q+2j)}{2j}(q+2j,q)$$

$$+ \frac{e}{\epsilon} \sum_{\substack{j=2,1,\cdots \\ j\neq 0}}^{-\infty} \sum_{\substack{j'=2,1,\cdots \\ j+j'\neq 0}}^{-\infty} \frac{(q,q+2j)(q+2j,q+2j+2j')}{2j(2j+2j')}$$

$$\times (q+2j+2j',q) + \cdots.$$

Successive contributions $\chi^{(0)}, \chi^{(1)}, \cdots$ of χ form a rapidly decreasing sequence provided that

$$\left(\frac{e}{\epsilon}\right)^{1/2} \frac{\chi_{q\pm 2}}{\chi_q} < 1,$$

i.e.,

$$\exp\left(\mp \frac{1}{e^{1/2}} \int^z \frac{dz}{v^{1/2}(z)}\right) < \left(\frac{\epsilon}{e}\right)^{1/2}. \tag{55}$$

This relation allows arbitrarily large values of z (since $\epsilon \to \infty$) but excludes the region around $z = -\frac{1}{2}$ [in view of the logarithmic term in (47)].

Thus, we now have two paris of large-h or ϵ asymptotic expansions of the eigenvalues and eigenfunctions of the logarithmic potential over (presumably) the entire range of the independent variable. We could even proceed to demonstrate that the two types of eigenfunctions we have derived are proportional to each other in their common region of validity. Such a verification would proceed along the lines of Refs. 7 and 9.

4. UNIFORMLY CONVERGENT SOLUTIONS AND THE SCATTERING PROBLEM OF THE CUT-OFF POTENTIAL

For the extension of our analysis to the cut-off potential which also permits scattering, it is useful and desirable to have yet another type of solution. For this reason we now derive an expansion which is uniformly convergent for finite values of ϵ or h. This type of solution has also been obtained by Gesztesy and Pittner.[8] We therefore skip the proof of its convergence, but investigate in more detail its relevance for the scattering problem.

Our starting point is Eq. (7) which we write

$$\frac{d^2\phi}{dz^2} - L^2\phi = \epsilon z e^{2z}\phi. \tag{56}$$

Setting

$$\phi = e^{\pm Lz} f(z) \tag{57}$$

we find that f satisfies the following equation,

$$\frac{d^2}{dz^2} f \pm 2L \frac{df}{dz} = \epsilon z e^{2z} f. \tag{58}$$

From now on we consider only the upper equation. The solution of the lower equation then follows by changing the sign of L. We solve (58) by iteration. Thus, if f_L is the solution of

$$\frac{d^2}{dz^2}f_L + 2L\frac{d}{dz}f_L = 0$$

we have

$$f_L = \text{constant} \quad \text{or} \quad f_L \propto e^{-2Lz}.$$

We take the first of these alternatives because the second form leads to $\phi = e^{-Lz}f$. Ignoring an overall constant, we take $f_L = 1$, and set

$$f = 1 + \sum_{i=1}^{\infty} \epsilon^i f_i, \tag{59}$$

where f_i is the solution of

$$\frac{d^2}{dz^2}f_i + 2L\frac{d}{dz}f_i = ze^{2z}f_{i-1}. \tag{60}$$

Solving for f_1, we set $f_1 = e^{2z}g_1(z)$ and then $g_1(z) = az + b$ where a and b are constants. Then

$$g_1(z) = \frac{1}{4(L+1)}\left(z - \frac{L+2}{2(L+1)}\right).$$

Proceeding in this way we obtain the solution

$$\begin{aligned}\phi &= e^{Lz}(1 + \epsilon f_1 + \epsilon^2 f_2 + \cdots)\\ &= e^{Lz}\Bigg[1 + \frac{\epsilon e^{2z}}{4(L+1)}\left(z - \frac{L+2}{2(L+1)}\right) + \frac{\epsilon^2 e^{4z}}{32(L+1)(L+2)}\\ &\quad\times\left(z^2 - \frac{(L+1)(L+4)+(L+2)^2}{2(L+1)(L+2)}z + \frac{(L+1)(L+4)^2+(L+4)(L+2)^2-2(L+1)(L+2)}{8(L+1)(L+2)^2}\right) + \cdots\Bigg].\end{aligned} \tag{61}$$

As pointed out earlier, a second solution is obtained by reversing the sign of L or replacing l by $-l-1$.

Rewriting (61) in terms of r, we obtain

$$\begin{aligned}\psi(r) &= r^{l+1}\exp\left(-\frac{\alpha}{\beta}(l+1)\right)\\ &\quad\times\left[1 + \beta r^2 g_1\left(l, \ln r - \frac{\alpha}{\beta}\right)\right.\\ &\quad\left. + \beta^2 r^4 g_2\left(l, \ln r - \frac{\alpha}{\beta}\right) + \cdots\right],\end{aligned} \tag{62}$$

where $g_1, g_2, \cdots$ are new functions of $\ln r$. The solution (62) is obviously the so-called regular solution.

We now look at the expansion (61) more closely and sum up the terms containing leading powers of z. Looking at higher terms which we do not reproduce here we see that the leading powers can be summed to a Bessel function of imaginary argument, i.e., (apart from an overall constant),

$$\begin{aligned}\psi_R(r) &= r^{l+1}\exp\left(-\frac{\alpha}{\beta}(l+1)\right)\\ &\quad\times\left(\frac{(l+\tfrac{1}{2})!I_{l+1/2}(\pm r(\beta z)^{1/2})}{[\pm(r/2)(\beta z)^{1/2}]^{l+1/2}} + \cdots\right).\end{aligned} \tag{63}$$

Now, $I_n(x)$ has the following asymptotic behavior,

$$I_n(x) \simeq \frac{e^x}{(2\pi x)^{1/2}}\left[1 + O\left(\frac{1}{x}\right)\right].$$

Considering only the dominant term of this expansion and the first term summed in (63), we have (inserting a constant N)

$$\begin{aligned}\psi(r) &\overset{r\to\infty}{\simeq} N\frac{\exp[-(\alpha/\beta)(l+1)](l+\tfrac{1}{2})!2^{l+1/2}}{(2\pi)^{1/2}(\pm\beta z)^{(l+1)/2}}\\ &\quad\times\exp[\pm r(-\alpha+\beta\ln r)^{1/2}],\end{aligned} \tag{64}$$

where $z = -\alpha/\beta + \ln r$.

Thus, if $k^2 \equiv 2\mu E/\hbar^2$, and so

$$\alpha = k^2 + \frac{2\mu g \ln r_0}{\hbar^2}, \tag{65}$$

the solution ψ has the asymptotic form

$$\frac{\exp(\pm r\{-k^2 + 2\mu g \ln r_0/\hbar^2 + \beta\ln r\}^{1/2})}{(-k^2 + 2\mu g \ln r_0/\hbar^2 + \beta\ln r)^{(l+1)/2}}. \tag{66}$$

This behavior shows that for $r\to\infty$ $\ln r$ dominates over the energy k^2, and we do not have scattering. Of course, if the potential is cut off at some distance R_0, the term $\ln r$ does not arise in the asymptotic form (66) and scattering is allowed. In fact requiring the wavefunction to be continuous at R_0, the S matrix is

$$\begin{aligned}S &= e^{-2ikR_0}\left(\frac{d\psi_R(R_0)}{dr} + ik\psi_R(R_0)\right)\\ &\quad\times\left(\frac{d\psi_R(R_0)}{dr} - ik\psi_R(R_0)\right)^{-1}\end{aligned} \tag{67}$$

where the subscript R indicates that the solution used is the regular solution discussed above. The Jost function $f_+(k) = f_-(e^{-i\pi}k)$ is given by

$$f_+(k) = e^{ikR_0}\left(\frac{d\psi_R(R_0)}{dr} - ik\psi_R(R_0)\right).$$

Its zeros determine the eigenvalues or—as was shown in the context of the linear power potential[10]—an expansion

$$q = q_0 + O\left(\frac{1}{R_0}\right),$$

where q_0 is exactly an odd integer, and q only approximately. For $R_0 \to \infty$, i.e., the potential with no cut off, $q \to q_0$. This is the case we have discussed previously.

The asymptotic form (63) suggests that solutions can be found in terms of cylindrical functions. This is in fact the case. Changing the independent variable of (56) to

$$\omega = [\epsilon z e^{2z}]^{1/2}$$

we obtain the equation

$$\frac{d^2\phi}{d\omega^2} + \frac{1}{\omega}\cdot\frac{d\phi}{d\omega} - \left(1 + \frac{L^2}{\omega^2}\right)\phi = -\left(1 + \frac{L^2}{\omega^2}\right)\frac{(1+4z)}{(1+2z)^2}\phi + \frac{2}{\omega}\cdot\frac{1}{(1+2z)^2}\cdot\frac{d\phi}{d\omega}. \tag{68}$$

For $|z| \to \infty$ the terms on the right-hand side of this equation become negligible and so

$$\phi \to I_L(\omega)$$

as we found above. We do not pursue this solution in further detail in the present context, although it would be interesting to know the complete iterative solution of (68).

5. UNIFORM ASYMPTOTIC SOLUTIONS

The last type of solution we wish to discuss complements the uniformly convergent solution for large values of ϵ, i.e., it is a uniform asymptotic expansion for large values of ϵ.

In our basic equation (56) we set

$$\phi = A(z)\exp\left(\int^z \epsilon^{1/2} z^{1/2} e^z\, dz\right). \tag{69}$$

Then $A(z)$ is found to satisfy the equation

$$2\frac{dA}{dz} + \left(1 + \frac{1}{2z}\right)A = \frac{1}{\epsilon^{1/2} z^{1/2} e^z}\left(L^2 A - \frac{d^2 A}{dz^2}\right). \tag{70}$$

For $\epsilon^{1/2} \to \infty$ we can neglect the terms on the right-hand side to a first approximation. Then A_0, the first term of the expansion

$$A = A_0 + \frac{1}{\epsilon^{1/2}}A_1 + \frac{1}{\epsilon}A_2 + \cdots \tag{71}$$

satisfies

$$2\frac{dA_0}{dz} + \left(1 + \frac{1}{2z}\right)A_0 = 0$$

so that

$$A_0 = Ce^{-z/2} z^{-1/4}, \tag{72}$$

C being an overall multiplicative constant. The equation for A_1 is

$$2\frac{dA_1}{dz} + \left(1 + \frac{1}{2z}\right)A_1 = \frac{A_0^3}{C^2}\left(L^2 - \frac{4z^2 + 4z + 5}{16z^2}\right) \tag{73}$$

which can be solved by first setting

$$A_1 = \frac{A_0^3}{C^2} f_1$$

where f_1 satisfies

$$\frac{df_1}{dz} - \left(1 + \frac{1}{2z}\right)f_1 = \tfrac{1}{2}L^2 - \frac{4z^2 + 4z + 5}{32z^2}.$$

This equation is again of a standard type but is nontrivial. Its solution can be given only in terms of the incomplete gamma function Γ,

$$\int_\infty^z z^{-n/2} e^{-z}\, dz = -\int_z^\infty z^{-n/2} e^{-z}\, dz = -\Gamma\left(1 - \frac{n}{2}, z\right). \tag{74}$$

Then

$$f_1 = z^{1/2} e^z\left[-\tfrac{1}{2}(L^2 - \tfrac{1}{4})\Gamma(\tfrac{1}{2}, z) + \tfrac{1}{8}\Gamma(-\tfrac{1}{2}, z) + \tfrac{5}{32}\Gamma(-\tfrac{3}{2}, z)\right].$$

Proceeding in this manner we find the solution

$$\phi = Ce^{-z/2} z^{-1/4} \exp\left[\epsilon^{1/2}\int^z z^{1/2} e^z\, dz\right] \times\left(1 + \frac{1}{\epsilon^{1/2}}\left[-\tfrac{1}{2}(L^2 - \tfrac{1}{4})\Gamma(\tfrac{1}{2}, z) + \tfrac{1}{8}\Gamma(-\tfrac{1}{2}, z) + \tfrac{5}{32}\Gamma(-\tfrac{3}{2}, z)\right] + O(1/\epsilon)\right). \tag{75}$$

This expansion is asymptotic in $\epsilon^{1/2}$ and presumably is valid over the entire range of z. We do not explore it in more detail in the present context.

Without going into details we mention that it is also possible to derive Fourier-types of solutions in rising powers of ϵ. These can be obtained by replacing the factor z or the right-hand side of (56) by its expansion in terms of $\sin nz$ (n an integer). This solution parallels the well known Fourier-type solution of the Mathieu equation.

6. The logarithmic potential and particle spectroscopy

We now investigate the implications of the logarithmic potential for the level spacing and leptonic widths of superheavy quark-antiquark bound states.

Solving the expansion (31) for l we obtain the Regge trajectories

$$l = \alpha_q(E) = -\tfrac{1}{2} + \frac{h^2}{(-)\,2^{3/2}}\left[1 - \frac{2q}{h^2} - \frac{69q^2 + 1}{2\cdot 3^2 h^4}\right.$$

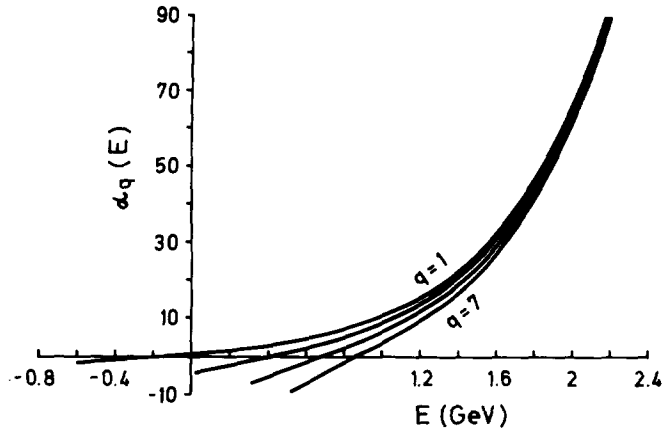

FIG. 1. Regge trajectories for the logarithmic potential for $r_0 = 5\ \text{GeV}^{-1}$, $g = 0.5692$ GeV, $\mu = 0.78\ \text{GeV}/c^2$, and $q = 1,3,5,7$.

$$- \frac{q(1691q^2+23)}{2^3 3^3 h^6} + O\left(\frac{1}{h^8}\right)\Bigg], \tag{76}$$

where $q = 2n+1$, $n = 0,1,2,\cdots$, and

$$h^2 = \frac{2r_0}{\hbar}(2\mu g)^{1/2}\exp\left(\frac{E}{g} - \frac{1}{2}\right). \tag{77}$$

In Figs. 1 and 2 we show the behavior of these trajectories for quark masses $m_c = 1.56\ \text{GeV}/c^2$, $m_b = 4.9\ \text{GeV}/c^2$. We observe in particular that the spacing between successive trajectories diminishes with increasing mass of the quarks.

Solving our expansion (31) for E we obtain

$$E \simeq \frac{q}{2} - \frac{1}{2}\ln\left(\frac{8\mu g r_0^2}{\hbar^2}\right) + g\ln\Bigg[2q + \Bigg(8(l+\tfrac{1}{2})^2 + \tfrac{1}{9}(69q^2+1) + \frac{q(35q^2-1)}{108\{2q + [8(l+\tfrac{1}{2})^2 + \tfrac{1}{9}(69q^2+1)]^{1/2}\}}\Bigg)^{1/2}\Bigg]. \tag{78}$$

TABLE I. Masses in the ψ and Υ families for $r_0 = 5\ \text{GeV}^{-1}$ and $g = 0.5692$ GeV (the first states require quark masses of 1.56 and 4.9 GeV/c^2; input values are underlined).

Masses (GeV/c^2) calculated by using (a) formula (78)	(b) numerical integration
ψ(3.095)	ψ(3.095)
ψ(3.685)	ψ(3.685)
ψ(3.973)	ψ(4.008)
ψ(4.163)	ψ(4.233)
ψ(4.304)	ψ(4.405)
Υ(9.40)	Υ(9.40)
Υ(9.99)	Υ(9.99)
Υ(10.28)	Υ(10.31)
Υ(10.47)	Υ(10.54)

Since the quark mass is twice the reduced mass μ, the mass M of a bound quark-antiquark pair is given by

$$M = 4\mu + E.$$

We observe that the level spacing $M_n - M_{n'}$ or $E_n - E_{n'}$ is independent of the reduced mass μ of the quarks and the range parameter r_0 of the potential. This difference is seen to depend only on the coupling constant g (and, of course, on the quantum number q). Thus the level spacing between the first two radial states fixes g uniquely, and a real test of the level spacing due to the logarithmic potential is possible only if more than two radial states are known. In Table I we show masses of the ψ and Υ families obtained with the help of (78). In order to demonstrate the usefulness of this formula, we compare these values with those obtained by Quigg and Rosner[2] by numerical integration. It should be observed that for the lowest S wave states ($q = 1$) the binding energies are still slightly negative.

Decay widths are well known to play an important role in exploring the origin of a newly found hadronic state. The leptonic and hadronic decay widths of a vector quark-antiquark bound state such as the ψ can be expressed in terms of the S wave bound state wavefunction at the origin. Thus

$$\Gamma(\psi \to l\bar{l}) = \frac{16\pi\alpha^2 e_Q^2}{m_\psi^2}\,|\Psi(0)|^2$$

and

$$\Gamma(\psi \to \text{hadrons}) = \tfrac{160}{81}(\pi^2 - 9)\frac{\alpha_s^3}{m_\psi^2}\,|\Psi(0)|^2. \tag{79}$$

Here α is the fine structure constant, α_s the strong coupling constant, and e_Q is the charge of the constituent quark of ψ.

In order to derive an expression for $|\Psi(0)|^2$ we proceed as follows. In the case of an S wave bound state, the wavefunction is related to the potential V via the following expression,

$$|\Psi(0)|^2 = \frac{\mu}{2\pi}\int \Psi^*(\mathbf{r})\frac{dV}{dr}\Psi(\mathbf{r})\,d\mathbf{r}. \tag{80}$$

This relation is obtained by dividing the S wave radial wave

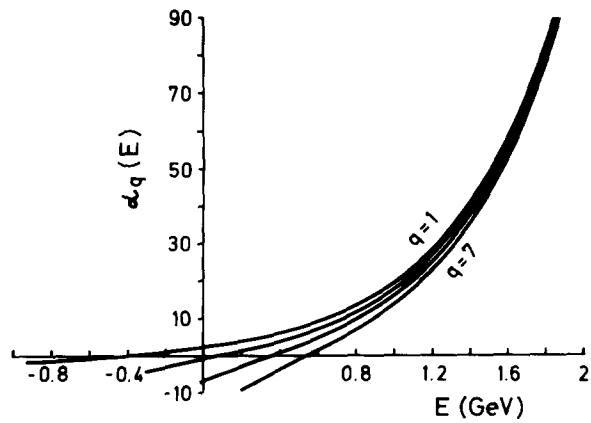

FIG. 2. Regge trajectories for the logarithmic potential for $r_0 = 5\ \text{GeV}^{-1}$, $g = 0.5692$ GeV, $\mu = 2.45\ \text{GeV}/c^2$, and $q = 1,3,5,7$.

equation (1) by ψ, differentiating with respect to r, taking the expectation value and performing two integrations by parts. Quigg and Rosner[6] have shown that in the case of a power potential the expression on the right-hand side of (80) can be related to the energy of the appropriate bound state. Their considerations do not hold for the logarithmic potential. We therefore proceed as follows. We assume that

$$V(r) = 0 \quad \text{for} \quad r < r_0,$$
$$= g \ln \frac{r}{r_0} \quad \text{for} \quad r > r_0. \tag{81}$$

This seems a reasonable restriction on $V(r)$ since it is well known that at short distances the Coulomb force (corresponding to exchange of massless gluons) is expected to become important. We ignore this component here.

For simplicity we restrict our considerations to the WKB approximation. The expression on the right-hand side of (80) can then be written

$$\frac{\mu}{2\pi} \int_0^{\bar{r}} dr\, \psi_{\mathrm{WKB}}^2(r) \frac{dV}{dr} \left(\int_0^{\bar{r}} dr\, \psi_{\mathrm{WKB}}^2(r) \right)^{-1}, \tag{82}$$

where $\bar{r}$ is the classical turning point given by

$$E - V(\bar{r}) = 0,$$

i.e.,

$$\bar{r} = r_0 e^{E/g}. \tag{83}$$

Now, the WKB approximated wavefunction ψ_{WKB} is proportional to $[E - V(r)]^{-1/4}$ multiplied by an oscillatory factor, the square of which can be taken as constant. Thus (82) can be written [using (81)]

$$\frac{\mu}{2\pi} \frac{\int_0^{\bar{r}} dr[E - V(r)]^{-1/2}(dV/dr)}{\int_0^{\bar{r}} dr[E - V(r)]^{-1/2}} = \frac{\mu}{2\pi} \frac{\int_{r_0}^{\bar{r}} dr[E - V(r)]^{-1/2}(dV/dr)}{\int_0^{\bar{r}} dr[E - V(r)]^{-1/2}}. \tag{84}$$

Setting $V(r) = E \sin^2\theta$, we have

$$\int_{r_0}^{\bar{r}} dr[E - V(r)]^{-1/2} \frac{dV}{dr} = 2E^{1/2}. \tag{85}$$

To evaluate the denominator of (84), we use the Bohr–Sommerfeld quantization relation (where $q = 2n + 1$, $n = 0,1,2,\cdots$)

$$\int_0^{\bar{r}} \{2\mu[E - V(r)]\}^{1/2} dr \simeq (n - \tfrac{1}{4})\pi.$$

Differentiating this expression with respect to n—the assumption of continuity in n implied by this procedure is approximately true and best when n is large—we have

$$\left(\frac{\mu}{2}\right)^{1/2} \frac{dE}{dn} \int_0^{\bar{r}} dr[E - V(r)]^{-1/2} = \pi. \tag{86}$$

Using (85) and (86) the expression (84) for $|\Psi(0)|^2$ becomes (to a reasonable approximation)

$$|\Psi(0)|^2 = \frac{\mu^{3/2}}{\pi^2} (2E)^{1/2} \frac{dE}{dq}. \tag{87}$$

TABLE II. The leptonic decay rates $\Gamma(V \to l^+l^-)$ in KeV (input values are underlined).

Mass (GeV/c^2)	Calculated	Experiment
ψ(3.095)	<u>4.80</u>	4.8 ± 0.6
ψ(3.685)	1.60	2.1 ± 0.3
ψ(3.973)	0.96	identification of level uncertain
ψ(4.163)	0.69	level not observed so far

Inserting (78) for E, this yields

$$|\Psi(0)|^2 \simeq 2\mu^{3/2} g^2 \left[1 - \ln\left(\frac{8\mu g r_0^2}{\hbar^2}\right)\right]^{1/2} \{q\pi^2[2 + \tfrac{1}{3}(69)^{1/2}]\}^{-1}. \tag{88}$$

We see, therefore, that for a logarithmic potential $q|\Psi(0)|^2$ is approximately constant, i.e., $|\Psi(0)|^2$ decreases with $q = 2n + 1$, $n = 0,1,2,\cdots$. This result should be compared with corresponding expressions for other types of potentials, since the dependence of the wavefunction on the radial quantum number has a strong bearing on the behavior of the leptonic decay rates of successive radial excitations such as ψ, ψ', ψ''. Thus, for the Coulomb potential we have[11]

$$|\Psi(0)|^2 \propto \frac{1}{n^3}$$

and for the generalized power potential[2,12] r^{λ}

$$|\Psi(0)|^2 \propto n^{2(\lambda-1)/(\lambda+2)}.$$

We see that for the logarithmic potential the decrease of the leptonic widths of sucessive radial excitations is considerably weaker than for the Coulomb potential, but is not independent of n as for the linear potential.

Using (88) we find that the logarithmic potential gives leptonic decay rates which are in reasonable agreement with data known for the ψ family, as Table II shows. However, hadronic decay rates come out too small, as the second of formulas (79) indicates, unless multichannel corrections are applied.

7. CONCLUSIONS

Our investigation shows that simple perturbation methods can be developed for solving explicitly the wave equation for a logarithmic potential. It is not difficult, but perhaps tedious, to include in the potential the necessary Coulomb component. In particular we derived two branches of asymptotic eigensolutions which are valid in different regions of r and together cover the entire range of the independent variable. That these solutions[13] are correct is verified by their respective eigenvalue expansions which are seen to be identical. Pairs of uniformly convergent and uniform asymptotic solutions have also been derived and are useful particularly for investigating the scattering problem of the cut off potential.

Our explicit eigenvalue expansion can be used directly for the calculation of binding energies, Regge trajetories, and leptonic decay rates of bound states of heavy quarks. It is

seen that the level spacing of radial and orbital excitations is independent of the quark masses, and that the leptonic decay rate of a radially excited state is inversely proportional to its radial quantum number. These results are in reasonable agrement with available data. However, two consecutive states do not suffice to test the predictions of the logarithmic potential. Hadronic decay rates may turn out to be too small. It can be instructive to keep this potential in mind when data of the new generation of accelerators become available.

Note added in proof: Finally we comment on the relation between the methods used in this paper and the WKB approximation. A reader of the preprint of this paper remarked that "a simple WKB approximation for the log r potential gives results which are more accurate than the expansions considered in this paper." This statement is wrong. The "simple WKB approximation" referred to by this reader is the dominant term of our expansion [i.e., the first two terms of expansion (31)], as can be seen by comparison with the work of Gesztesy and Pittner.[8] In fact, the expansions derived in this paper *are* the WKB expansions. In the present context we do not demonstrate this equivalence, but this has been done in the literature for the case of the Yukawa potential. Perturbation expansions similar to those derived here have been derived for the eigenvalues and eigenfunctions of the Yukawa potential in papers quoted under Ref. 14. The approximations derived for the same potential by the usual WKB method have been obtained by Boukema.[15] The first paper of this author discusses the usual WKB approximation and agreement with our expansion is obtained only up to the second term in the expansions. However, in the second paper a correction to the WKB approximation is investigated and complete agreement with our expansion is obtained (up to the first four terms calculated). Thus it may be argued that our methods are superior to the usual WKB approximation.

[1]L.M. Lederman, *Proceedings of the* 1977 *International Symposium on Lepton and Photon Interactions at High Energies*, edited by F. Gutbrod, DESY, Hamburg, 1977, p. 567.

[2]C. Quigg and J.L. Rosner, Phys. Lett. B **71**, 153 (1977).

[3]C. Quigg and J.L. Rosner, Phys. Lett. B **72**, 462 (1978).

[4]H.B. Thacker, C. Quigg, and J.L. Rosner, Phys. Lett. B **74**, 350 (1978).

[5]For the generalized power potential the eigenenergies and Regge trajectories have been calculated by S.K. Bose, A. Jabs, and H.J.W. Müller-Kirsten, Phys. Rev. D **13**, 1489 (1976).

[6]C. Quigg and J.L. Rosner, Fermilab-Pub-77/106-THY, 1977, and Fermilab-Pub-78/19-THY, 1978 (unpublished).

[7]For the Mathieu equation see, for instance, R.B. Dingle and H.J.W. Müller, J.F. Reine U. Angew. Math. **221**, 11 (1962) and H.A. Aly, H.J.W. Müller-Kirsten, and N. Vahedi-Fardi, J. Math. Phys. **16**, 961 (1975). However, the Mathieu equation is exceptionally simple. Our procedure here therefore parallels that developed for the more complicated case of the Gauss potential, H.J.W. Müller, J. Math. Phys. **11**, 355 (1970).

[8]F. Gesztesy and L. Pittner, University of Graz preprint UTP-04/77, 1977, to appear in J. Phys. A. These authors investigate the logarithmic potential with a view for application to the calculation of electron interferences observed in "Fresnel biprisma experiments" in which electrons are used instead of light and suitable electric fields instead of the optical lens.

[9]The same procedure has been applied previously to several other cases, such as the Schrödinger equation for a Gauss potential [H.J.W. Müller, J. Math. Phys. **11**, 355 (1970)] and the ellipsoidal wave equation [H.J.W. Müller, Math. Nachr. **31**, 89 (1966)].

[10]H.J.W. Müller-Kirsten and N. Vahedi, Phys. Rev. D **14**, 3079 (1976).

[11]L.D. Landau and E.M. Lifshitz, *Relativistic Quantum Theory* (Pergamon, New York, 1971), p. 288.

[12]H.J.W. Müller-Kirsten, G.E. Hite, and S.K. Bose, "Explicit solution of the wave equation for arbitrary power potentials with application to charmonium spectroscopy," University of Kaiserslautern preprint, 1978, unpublished.

[13]For the explicit normalization of these wavefunctions see R.S. Kaushal and H.J.W. Müller-Kirsten, "Explicit normalization of boundstate wavefunctions and the calculation of decay widths," University of Kaiserslautern preprint, 1979, unpublished.

[14]H.J.W. Müller, Ann. Phys. (Leipz.) **15**, 395 (1965); Physica **31**, 688 (1965); H.J.W. Müller and K. Schilcher, J. Math. Phys. **9**, 255 (1968); H.J.W. Müller-Kirsten and N. Vahedi-Faridi, J. Math. Phys. **14**, 1291 (1973).

[15]J. Boukema, Physica **30**, 1320, 1909 (1964).

The complete exact solution to the translation-invariant N-body harmonic oscillator problem[a)]

Richard L. Hall
Department of Mathematics, Concordia University, Montreal, Quebec, Canada

B. Schwesinger
Gesamthochschule Siegen, Fachbereich 7, Adolf Reichwein Strasse, 5900 Siegen 21, Germany

(Received 24 April 1979; accepted for publication 6 August 1979)

It is shown that Schrödinger's nonrelativistic equation for a translation-invariant system of N particles with arbitrary masses interacting by harmonic pair potentials with arbitrary coupling constants is exactly soluble. An explicit matrix KV is given whose eigenvalues and eigenvectors determine all the exact energies and corresponding eigenfunctions of the N-particle problem. The result is extended to include systems composed of an arbitrary number of groups of identical particles.

I. INTRODUCTION

We consider a system of N particles which interact by harmonic pair potentials and obey nonrelativistic quantum mechanics. The translation-invariant Hamiltonian H for this system may be written:

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^{\,2}}{2m_i} - \frac{1}{2m}\Big(\sum_{i=1}^{N} \mathbf{p}_i\Big)^2 + \sum_{\substack{i,j=1 \\ i<j}}^{N} g_{ij}(\mathbf{r}_i - \mathbf{r}_j)^2, \tag{1}$$

where $m = \sum_{i=1}^{N} m_i$ is the total mass and the coupling constants g_{ij} satisfy $g_{ij} = g_{ji} \geqslant 0$. In earlier articles (Hall, 1–3 Schwesinger[4,5]) we have constructed exact solutions to Schrödinger's equation for this problem in cases where the number of distinct masses and coupling constants is not too large. In the present article we demonstrate that the general problem with positive or zero coupling constants (this assumption can be weakened, as we show in Sec. II) is always exactly soluble. The result is, of course, not surprising from the physical point of view. The mathematical difficulty is to diagonalize simultaneously both the potential and kinetic energy expressions in terms of a suitable set of *translation-invariant* relative coordinates. We shall do this by first expressing H in terms of a simple set of relative coordinates and then showing how a new set can always be found which diagonalizes the general problem. In Sec. III we extend the result of systems composed of groups of identical particles.

II. THE MAIN RESULT

We denote the column vectors of the original position and momentum coordinates by $\mathbf{r}$ and $\mathbf{p}$ respectively and we define new coordinates $\boldsymbol{\sigma}' = B\mathbf{r}$ and momenta $\boldsymbol{\tau}' = (B^{-1})^T\mathbf{p}$ by the matrix

$$B = \begin{bmatrix} \frac{m_1}{m} & \frac{m_2}{m} \cdots & \frac{m_N}{m} \\ 1 & 0 \cdots 0 & -1 \\ 0 & 10 \cdot 0 & -1 \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots 1 & -1 \end{bmatrix}. \tag{2}$$

The unprimed variables $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ represent, respectively, column vectors of the *relative* coordinates and momenta, i.e., $\boldsymbol{\sigma} = (\sigma_2, \sigma_3, \ldots, \sigma_N)^T$ and $\boldsymbol{\tau} = (\tau_2, \tau_3, \ldots, \tau_N)^T$. If by M we denote the diagonal matrix $[m_i]$ of the masses, then the *total* kinetic energy may by written:

$$\text{K.E.} = \tfrac{1}{2}\,\mathbf{p}^T M^{-1}\mathbf{p} = \tfrac{1}{2}(\boldsymbol{\tau}')^T B M^{-1} B^T \boldsymbol{\tau}',$$

and, in view of Eq. (2), we have

$$\text{K.E.} = \tfrac{1}{2}(\boldsymbol{\tau}')^T \begin{bmatrix} \frac{1}{m} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & K & \\ 0 & & & \end{bmatrix} \boldsymbol{\tau}' \tag{3}$$

where the $(N-1) \times (N-1)$ matrix K is invertible and positive definite since M has these properties. The matrix K has the general form

[a)]This work was supported in part by a Natural Sciences and Engineering Council Canada Grant No. A3438.

 0022-2488/79/122481-03$01.00

$$K = \begin{bmatrix} \left(\frac{1}{m_1}+\frac{1}{m_N}\right) & \frac{1}{m_N} & \frac{1}{m_N}\cdots & \frac{1}{m_N} \\ \frac{1}{m_N} & \left(\frac{1}{m_2}+\frac{1}{m_N}\right) & \frac{1}{m_N}\cdots & \frac{1}{m_N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{1}{m_N} & \frac{1}{m_N} & \frac{1}{m_N}\cdots & \left(\frac{1}{m_{N-1}}+\frac{1}{m_N}\right) \end{bmatrix}. \tag{4}$$

Thus the *relative* kinetic energy becomes

$$\text{Relative K.E.} = \tfrac{1}{2}\boldsymbol{\tau}^T K \boldsymbol{\tau}. \tag{5}$$

In term of the relative coordinates σ the potential energy operator in H may be written

$$\text{P.E.} = \sum_{i<j<N} g_{ij}(\boldsymbol{\sigma}_{i+1} - \boldsymbol{\sigma}_{j+1})^2 + \sum_{i<N} g_{iN}\boldsymbol{\sigma}_{i+1}^2,$$

i.e.,

$$\text{P.E.} = \boldsymbol{\sigma}^T V \boldsymbol{\sigma}, \tag{6}$$

where the positive definite matrix V has the general form

$$V = \begin{bmatrix} \left(\sum_{j\neq 1} g_{1j}\right) & -g_{12} & -g_{13}\cdots & -g_{1N-1} \\ -g_{21} & \left(\sum_{j\neq 2} g_{2j}\right) & -g_{23}\cdots & -g_{2N-1} \\ \cdot & & & \\ \cdot & & & \\ -g_{N-11} & -g_{N-12} & \cdots & \left(\sum_{j\neq N-1} g_{N-1j}\right) \end{bmatrix}. \tag{7}$$

The positive definiteness of V (which is certainly guaranteed by our assumption $g_{ij} \geqslant 0$) is not required by our diagonalization proof: This property of V is necessary for a physically meaningful solution to the problem and, as we have seen in earlier examples (Hall[3]), this latter condition may be satisfied in special cases even when some of the g_{ij} are *negative*.

Hence, from Eqs. (1), (5), and (6) we have

$$H = \tfrac{1}{2}\boldsymbol{\tau}^T K \boldsymbol{\tau} + \boldsymbol{\sigma}^T V \boldsymbol{\sigma}. \tag{8}$$

If we now choose a new set of *relative* coordinates by means of the invertible $(N-1)\times(N-1)$ matrix R, i.e.,

$$\boldsymbol{\rho} = R^{-1}\boldsymbol{\sigma} \quad \text{and} \quad \boldsymbol{\pi} = (R)^T\boldsymbol{\tau}, \tag{9}$$

then the Hamiltonian may be written from Eqs.(8) and (9) in the form

$$H = \tfrac{1}{2}\boldsymbol{\pi}^T(R^T K^{-1} R)^{-1}\boldsymbol{\pi} + \boldsymbol{\rho}^T R^T V R \boldsymbol{\rho}. \tag{10}$$

Since K^{-1} and V are both symmetric and positive definite, we know, for example from Noble (Ref. 6, Theorem 12.7, p. 396), that a matrix R exists so that

$$R^T K^{-1} R = I \quad \text{and} \quad R^T V R = \Lambda, \tag{11}$$

where $\Lambda = [\lambda_i]$ is diagonal and $\lambda_i \geqslant 0$. The *exact* eigenvalues of H are therefore given by

$$E = \sum_{i=2}^{N} \{2(a_i + b_i + c_i) + 3\}\hbar\lambda_i^{1/2}/\sqrt{2}, \tag{12}$$

where a_i, b_i, and c_i are arbitrary positive or zero integers. From (11) we see that the λ_i are given by the $(N-1)$th degree polynomial equation

$$|KV - \lambda I| = 0. \tag{13}$$

The columns $\mathbf{x}_i (i = 2,3,...,N)$ of R are the eigenvectors of the matrix KV associated with the corresponding eigenvalue λ_i, i.e.,

$$KV\mathbf{x}_i = \lambda_i \mathbf{x}_i. \tag{14}$$

Thus if we find the λ_i from Eq. (13), we may then construct the matrix R with the vectors $\mathbf{x}_i$ obtained from Eq. (14). The relative coordinates $\boldsymbol{\rho}$ are given in terms of the $(N-1)$ pair-distance coordinates $\boldsymbol{\sigma}$ by Eq. 9, i.e., $\boldsymbol{\rho} = R^{-1}\boldsymbol{\sigma}$. The exact eigenfunctions of H corresponding to the eigenvalues E [Eq. (12)] are then the appropriate products of $3(N-1)$ Hermite functions in the $3(N-1)$ variables $(\boldsymbol{\rho}_2, \boldsymbol{\rho}_3, ..., \boldsymbol{\rho}_N)$.

III. SYSTEMS COMPOSED OF GROUPS OF IDENTICAL PARTICLES

We consider S groups of identical particles with N_α particles of mass m_α and $\alpha = 1,2,...,S$. The translation-invariant Hamiltonian H for the entire system of $N = \sum_{\alpha=1}^{S} N_\alpha$ particles is as follows:

$$H = \sum_{\alpha=1}^{S} \left\{ \sum_{i_\alpha=1}^{N_\alpha} \frac{\mathbf{p}_{i_\alpha}^2}{2m_\alpha} + \sum_{i_\alpha<j_\alpha}^{N_\alpha} g_\alpha (\mathbf{r}_{i_\alpha} - \mathbf{r}_{j_\alpha})^2 \right\} + \sum_{\alpha<\beta}^{S} \sum_{i_\alpha, j_\beta}^{N_\alpha, N_\beta} g_{\alpha\beta} (\mathbf{r}_{i_\alpha} - \mathbf{r}_{j_\beta})^2 - \left(\sum_{\alpha, i_\alpha} \mathbf{p}_{i_\alpha} \right)^2 \left(2 \sum_{\alpha} N_\alpha m_\alpha \right)^{-1}. \tag{15}$$

For each group α of identical particles we define $(N_\alpha - 1)$ Jacobi orthogonal relative coordinates $\boldsymbol{\rho}_{i_\alpha}$ (as in Ref. 3) with $i_\alpha = 2,3,...,N_\alpha$, a *group* center-of-mass coordinate $\mathbf{R}_\alpha = (1/N_\alpha)\sum_{i_\alpha}^{N_\alpha} \mathbf{r}_{i_\alpha}$, and also the corresponding conjugate momenta $\boldsymbol{\pi}_{i_\alpha}$ and $\mathbf{P}_\alpha$. By using the operator methods of Refs. 4 and 5 or by a direct calculation we find that Eq. (15) can be written in the following separated form:

$$H = \sum_{\alpha=1}^{S} H_\alpha + H_c, \tag{16}$$

where

$$H_\alpha = \sum_{i_\alpha=2}^{N_\alpha} \left\{ \frac{\boldsymbol{\pi}_{i_\alpha}^2}{2m_\alpha} + (N_\alpha g_\alpha + h_\alpha) \boldsymbol{\rho}_{i_\alpha}^2 \right\},$$

$$H_c = \sum_{\alpha=1}^{S} \frac{\mathbf{P}_\alpha}{2M_\alpha} - \left(\sum_{\alpha=1}^{N_\alpha} \mathbf{P}_\alpha \right)^2 \left(2 \sum_{\alpha} M_\alpha \right)^{-1} + \sum_{\alpha<\beta} N_\alpha N_\beta g_{\alpha\beta} (\mathbf{R}_\alpha - \mathbf{R}_\beta)^2,$$

and

$$M_\alpha = N_\alpha m_\alpha, \quad h_\alpha = \sum_{\beta=1}^{S} g_{\alpha\beta} N_\beta, \quad g_{\alpha\alpha} = 0.$$

Since the H_α are diagonal and since H_c has exactly the form of the Hamiltonian which we diagonalized in Sec. II. above, Eq. (16) provides the complete solution of the problem for S groups of identical particles. In Ref. 3 we solved the special case $S = 2$ but there is an error in Eqs. (3) and (7) of that reference: The inter-center-of-mass kinetic energy term $(1/N)(N_2/m_1 + N_1/m_2)\boldsymbol{\pi}^2$ should be divided by 2; with this correction to the earlier work it becomes consistent with the general result, Eq. (16) above. For each α the variables $\mathbf{R}_\alpha$ and $\mathbf{P}_\alpha$ are symmetric under the permutation of the individual-particle indices of the corresponding set of N_α identical particles so that the burden of spatial permutation symmetry is carried entirely by the eigenfunctions of H_α (for bosons and fermions the details are given in Refs. 3 and 4).

IV. CONCLUSION

We have transformed the translation-invariant Hamiltonian for an arbitrary system of particles interacting by harmonic pair potentials into diagonal form. This provides exact solutions for a large collection of models suitable for molecular physics and for systems of quarks.[7,8] These solutions also allow energy lower bounds (e.g., Ref. 9) for general interactions to be compared with the exact energies in the special case of harmonic potentials.

ACKNOWLEDGMENT

One of us (R.L.H.) would like to thank Professors E.J. Burge and H.R. Post for the facilities provided at Chelsea College, London, where some of this work was carried out.

[1] R.L. Hall, Phys. Lett. A **65**, 85 (1978).
[2] R.L. Hall, J. Phys. A: Math. Gen. **11**, 1227 (1978).
[3] R.L. Hall, J. Phys. A: Math. Gen. **11**, 1235 (1978).
[4] B. Schwesinger, Z. Phys. A **282**, 173 (1977).
[5] B. Schwesinger, Z. Phys. A **287**, 373 (1978).
[6] B. Noble, *Applied Linear Algebra* (Prentice-Hall, Englewood Cliffs, N.J. 1969).
[7] R.P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971).
[8] N. Isgur and G. Karl, Phys. Lett. B **74**, 353 (1978).
[9] R.L. Hall, J. Math. Phys. **19**, 1969 (1978).

The radial reduced Coulomb Green's function

Bruce R. Johnson and Joseph O. Hirschfelder
Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706

(Received 18 July 1979; accepted for publication 24 August 1979)

The reduced Green's function, $g_{nl}(r,r')$, of the radial hydrogenic Schrödinger equation is simplified for all values of n and l, $l \leq n-1$, to a closed form appropriate for analytic treatments of Rayleigh–Schrödinger perturbation theory. Integral moments of the form $\int_0^\infty dr'\, g_{nl}(r,r')(r')^{k+2}\exp(-Zr'/n)$ are given. It is also shown how g_{nl} is connected to $g_{n,l\pm1}$ by the ladder operators of the factorization method. Recursion relations are derived between integrals that arise in perturbation theory. The above results are generalized to the case $l \geq n$, which occurs in the Green's functions required for the Rayleigh–Schrödinger perturbation treatment even though it does not arise for the eigenfunctions. As an example of the use of the reduced Green's functions, the first-order wave function and second-order energy corresponding to the spin-orbit interaction is evaluated for any bound state.

I. INTRODUCTION

A. Outline

Rayleigh–Schrödinger perturbation theory (RSPT) for hydrogenic systems can easily be performed analytically by use of the reduced Coulomb Green's function (RCGF).[1-8] The three-dimensional RCGF can be expanded in terms of spherical harmonics and the radial RCGF's, $g_{nl}(r,r')$; g_{nl}, in contrast to the bound state radial wave functions, is defined for both $l \leqslant n-1$ and $l \geqslant n$. The expression for $g_{nl}(l \leqslant n-1)$ given by Sherstyuk[4] contains a confluent hypergeomtric series which we have summed to closed form, permitting us to present simple expressions for all of the radial RCGF's needed in bound-state RSPT, i.e., for both $l \leqslant n-1$ and $l \geqslant n$. We have used these to calculate radial integrals of the form

$$\int_0^\infty dr'\, g_{nl}(r,r')(r')^{k+2}e^{-Zr'/n} \tag{1.1}$$

for arbitrary l and n, $k \geqslant -l-2$. These integrals may be of use in evaluating first-order wave functions.

The ladder operators of the factorization method[9] are shown to connect g_{nl} to $g_{n,l\pm1}$ in a simple manner, and to generate differential recursion relations between integrals such as

$$\int_0^\infty dr'\, g_{nl}(r,r')(r')^{k+2}R_{nl'}(r'), \tag{1.2}$$

where $R_{nl'}$ is a radial wave function. A nondifferential recursion relation which should be useful in applications is also derived by a procedure previously used to relate radial matrix elements.[10]

In Sec. II we present a brief description of the radial RCGF's followed by a summary of the above-mentioned results. For the benefit of the reader not concerned with derivations, these are delayed until the latter sections of the paper.

Equation (2.11) is the usual eigenfunction expansion of g_{nl} and demonstrates how, for both $l \leqslant n-1$ and $l \geqslant n$, g_{nl} is constructed from radial continuum functions labeled by the quantum number l and from the bound radial functions $R_{n'l}$, $n' \geqslant l+1$. The eigenfunction expansion shows simply how the case $l \geqslant n$ appears for the RCGF's.

Equation (2.15) is the closed form of the hypergeometric series appearing in Sherstyuk's work, while the proof is deferred to Sec. IV. Equations (2.13) and (2.18) are the relatively simple final forms of g_{nl} for $l \geqslant n$ and $l \leqslant n-1$, respectively.

The integral moments (1.1) are tabulated in Eqs. (2.23)–(2.28). No proof is given here.

Equations (2.35) and (2.36) demonstrate how the same ladder operators that relate R_{nl} to $R_{n,l\pm1}$ may be used analogously for the radial RCGF's while Eqs. (2.39) and (2.40) give examples of the sort of differential recursion relations the ladder operators generate between integrals of the form (1.2). Derivations of those results can be found in Sec. V.

Equation (2.49) is the nondifferential relation between these integrals, and in Sec. VI we show the steps leading up to it. We present in Sec. III an application of these results in calculating perturbed wave functions, adding spin to the hydrogenic model and taking as our perturbation the spin–orbit interaction. Finally, in Appendix B we give an alternative means of obtaining the RCGF's considered by Sherstyuk.

B. Relationship of the present treatment to previous research

The function g_{nl} for $l \leqslant n-1$ has been evaluated in closed form previously in special cases by Hameka[1] ($n=1$), Hostler[7] ($l=0$, n arbitrary), and Laurenzi and Flamberg[11] ($n=1,2,3$). The latter authors managed to sum Sherstyuk's hypergeometric series for the first few quantum numbers, leading us to consider the general case. Their work has pinpointed one of the difficulties that still exists in evaluating integrals over g_{nl}, however: A large amount of algebra is involved which, while straightforward, is also tedious. The integrals and recursion relations we present for both $l \leqslant n-1$ and $l \geqslant n$ are efforts to ameliorate this situation.

It should be noted that g_{nl} has been given before as an infinite expansion in the Sturmian representation.[12,13] These forms are actually easier to use for some calculations[12-15] and have the merit of unifying the two cases $l \leqslant n-1$ and $l \geqslant n$. In general, however, it is desirable to have g_{nl} in closed form.

Throughout the paper, we have limited our scope to

 0022-2488/79/122484-18$01.00

single integrals that would be encountered in finding first-order wave functions for perturbing potentials in which the radial dependence consists of positive or negative powers of r. [The perturbation $r^k\exp(-\beta r)$ has been treated by Sherstyuk.[13]] The simple recursion relations between these integrals suggest that a more detailed study may be warranted, especially since the hydrogenic system has proven such a valuable testing ground for various model theories. We expect that the results given here can be extended to use in higher order calculations. It also seems likely that the treatments of Secs. V and VI (and Appendix B) can be made for other types of physical systems.

II. RESULTS

A. Review and description of the Green's functions

The three-dimensional Schödinger equation for a hydrogenic ion with nuclear charge Z is[16]

$$\left[-\tfrac{1}{2}\nabla^2-\frac{Z}{r}-E\right]\psi(\mathbf{r})=0\,. \tag{2.1}$$

The bound states of this equation are described by the energy eigenvalues $E_n=-(Z^2/2n^2)$, $n=1,2,\ldots$, and the orthonormal eigenfunctions ($l=0,1,\ldots,n-1$; $m=-l,-l+1,\ldots,l-1,l$)

$$\psi_{nlm}(\mathbf{r})=R_{nl}(r)Y_{lm}(\theta,\varphi)\,, \tag{2.2}$$

where

$$R_{nl}(r)=\left[\frac{Z^3(n-l-1)!}{(n+l)!}\right]^{1/2}\frac{2}{n^2}x^l L_{n-l-1}^{2l+1}(x)e^{-x/2} \tag{2.3}$$

and

$$x=\frac{2Zr}{n},\quad L_{n-l-1}^{2l+1}(x)=\sum_{k=0}^{n-l-1}\binom{n+l}{n-l-1-k}\frac{(-x)^k}{k!}\,.$$

In the above, L_{n-l-1}^{2l+1} is an associated Laguerre polynomial[17] and Y_{lm} is a spherical harmonic. The complete set of states is well known to contain continuum states for which $E>0$.[18]

The Green's function $G(\mathbf{r},\mathbf{r}';E)$ associated with Eq. (2.1) satisfies (for E not an eigenvalue)

$$\left[-\tfrac{1}{2}\nabla^2-\frac{Z}{r}-E\right]G(\mathbf{r},\mathbf{r}';E)=-\delta(\mathbf{r}-\mathbf{r}')\,. \tag{2.4}$$

The eigenfunction expansion of G is given by[19]

$$\begin{aligned}G(\mathbf{r},\mathbf{r}';E)&=\sum_{n'=1}^{\infty}\sum_{l=0}^{n'-1}\sum_{m=-l}^{l}\frac{\psi_{n'lm}(\mathbf{r})\psi^*_{n'lm}(\mathbf{r}')}{E-E_{n'}}\\&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}Y_{lm}(\theta,\varphi)Y^*_{lm}(\theta',\varphi')\\&\quad\times\sum_{n'=l+1}^{\infty}\frac{R_{n'l}(r)R^*_{n'l}(r')}{E-E_{n'}}\\&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}Y_{lm}(\theta,\varphi)Y^*_{lm}(\theta',\varphi')\\&\quad\times g_l(r,r';E)\,.\end{aligned} \tag{2.5}$$

The radial Coulomb Green's function g_l satisfies the inhomogeneous equation

$$[H_l-E\,]g_l(r,r';E)=-\frac{\delta(r-r')}{rr'}\,, \tag{2.6}$$

where

$$H_l=-\frac{1}{2r}\frac{\partial^2}{\partial r^2}r+\frac{l(l+1)}{2r^2}-\frac{Z}{r}\,. \tag{2.7}$$

Combined with regularity conditions as r or r' tend to zero or infinity, g_l is given by[19,20]

$$\begin{aligned}g_l(r,r';E)=&-\frac{4Z}{\nu}\frac{\Gamma(l+1-\nu)}{\rho_<\rho_>}\mathcal{M}_{\nu,l+1/2}(\rho_<)\\&\times W_{\nu,l+1/2}(\rho_>)\,,\end{aligned} \tag{2.8}$$

with

$$\rho=2Zr/\nu,\quad \rho'=2Zr'/\nu\,, \tag{2.9}$$

and

$$\nu=\sqrt{-\frac{Z^2}{2E}}\,. \tag{2.10}$$

$\mathcal{M}_{\nu,l+1/2}$ and $W_{\nu,l+1/2}$ are Whittaker functions,[21] $\Gamma(l+1-\nu)$ is the gamma function, and $\rho_<$ ($\rho_>$) is the lesser (greater) of ρ and ρ'.

For applications in bound state RSPT,[22] we actually desire the RCGF, given by

$$\begin{aligned}G_n(\mathbf{r},\mathbf{r}')&=\sum_{\substack{n'=1\\n'\neq n}}^{n'-1}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{\psi_{n'lm}(\mathbf{r})\psi^*_{n'lm}(\mathbf{r}')}{E_n-E_{n'}}\\&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}Y_{lm}(\theta,\varphi)Y^*_{lm}(\theta',\varphi')\\&\quad\times\sum_{\substack{n'=l+1\\n'\neq n}}^{\infty}\frac{R_{n'l}(r)R^*_{n'l}(r')}{E_n-E_{n'}}\\&=\sum_{l,m}Y_{lm}(\theta,\varphi)Y^*_{lm}(\theta',\varphi')g_{nl}(r,r')\,,\end{aligned} \tag{2.11}$$

where $g_{nl}(r,r')$ is the radial RCGF, orthogonal to R_{nl}. Since the summation over n' in g_{nl} starts at $n'=l+1$, there are two distinct cases: For $l\leqslant n-1$, the term with $n'=n$ must be explicitly excluded and, for $l\geqslant n$, no such term would arise anyway, making the restriction $n'\neq n$ superfluous.

Thus, G_n and g_{nl} may be obtained from G and g_l, respectively, by setting $E=E_n$, the energy of the nth bound level, after removing any pole terms. Hostler[3] has given a prescription for this process and $G_n(\mathbf{r},\mathbf{r}')$ has been found in closed form.[5,6,8] For the radial functions, this prescription reads as

$$\begin{aligned}g_{nl}(r,r')&=\left[\frac{\partial}{\partial E}(E-E_n)g_l(r,r';E)\right]_{E=E_n}\\&=\frac{n}{2}\left[\frac{\partial}{\partial(1/\nu)}\left(\frac{1}{\nu^2}-\frac{1}{n^2}\right)g_l\left(r,r';-\frac{Z^2}{2\nu^2}\right)\right]_{\nu=n}\,.\end{aligned} \tag{2.12}$$

It can be easily verified from the series expansion of g_{nl} that Eq. (2.12) is valid for both of the cases $l\leqslant n-1$ and $l\geqslant n$, even though we can obtain g_{nl} in closed form in the second case by simply setting $\nu=n$.[1] In this case,[11,23]

$$g_{nl}(r,r') = \lim_{\nu \to n} g_l\left(r,r';-\frac{Z^2}{2\nu^2}\right) = -\frac{4Z}{n}\frac{(l-n)!}{x_< x_>}\mathscr{M}_{n,l+1/2}(x_<)W_{n,l+1/2}(x_>)$$
$$= (-)^{l+1-n}\frac{4Z}{n}(l-n)!(l+n)!(xx')^{-l-1}e^{-(x+x')/2}\left[\sum_{i=0}^{l+n}\binom{2l-i}{l-n}\frac{x_>^i}{i!}\right]$$
$$\times\left[e^x\sum_{j=0}^{l-n}\binom{2l-j}{l+n}\frac{(-x_<)^j}{j!} - \sum_{k=0}^{l+n}\binom{2l-k}{l-n}\frac{x_<^k}{k!}\right], \tag{2.13}$$

where

$$x = 2Zr/n, \quad x' = 2Zr'/n.$$

For $l \leqslant n-1$, the details are more complex. Using Eq. (2.12), Sherstyuk[4] reduced g_{nl} to the form[24]

$$g_{nl}(r,r') = \frac{4Z}{n}\frac{p!}{q!}(xx')^l e^{-(x+x')/2}\left\{L_p^s(x)L_p^s(x')\left[\log x + \frac{x+x'}{2n} - \psi(p+1) - \frac{2s+3}{2n}\right] + \frac{x}{n}L_{p-1}^{s+1}(x)L_p^s(x')\right.$$
$$+ \frac{x'}{n}L_{p-1}^{s+1}(x')L_p^s(x) + L_p^s(x_<)\left[\sum_{k=0}^{p-1}A_k x_>^k - \sum_{k=1}^{s}\binom{q}{p+k}\frac{(k-1)!}{x_>^k}\right]$$
$$\left. + L_p^s(x_>)\sum_{k=1}^{p}B_k x_<^k + L_p^s(x_>)\frac{(-x_<)^{p+1}}{(p+1)!(q+1)}{}_2F_2(1,1;q+2,p+2;x_<)\right\}. \tag{2.14}$$

In this expression we use the definitions[25]

$$p = n-l-1,\ s = 2l+1,\ q = n+l,\ A_k = \frac{(-)^k}{k!}\binom{q}{p-k}\sum_{j=k+1}^{p}\frac{2j+s}{j(j+s)},\ B_k = \frac{(-)^k}{k!}\binom{q}{p-k}\sum_{j=0}^{k-1}\frac{1}{p-j}.$$

The function $\psi(p+1)$ is the logarithmic derivative of the gamma funciton,[26]

$$\psi(p+1) = -\gamma + \sum_{k=1}^{p}\frac{1}{k},\ \gamma = 0.57721\ldots \text{ (Euler's constant)},$$

and the ${}_2F_2$ function is a confluent hypergeometric series.[27]

As noted in the introduction, Laurenzi and Flamberg[11] have evaluated this series in the cases $n = 1,2$, and 3. Proceeding along their lines, we found that this is possible for the general case (see Sec. IV). In the notation of Eq. (2.14), we obtain

$$\frac{(-x)^{p+1}}{(p+1)!(q+1)}{}_2F_2(1,1;q+2,p+2;x) = e^x\left[\sum_{k=1}^{s}\binom{q-k}{p}\frac{(k-1)!}{x^k} + \Phi_{nl}(x)\right]$$
$$- \sum_{k=1}^{s}\binom{q}{p+k}\frac{(k-1)!}{x^k} + L_p^s(x)[\log x - Ei(x) - \psi(q+1)]$$
$$+ \sum_{k=0}^{p-1}A_k x^k - \sum_{k=1}^{p}B_k x^k, \tag{2.15}$$

where $Ei(x)$ is the exponential integral[28] and

$$\Phi_{nl}(x) = \sum_{\lambda=1}^{p}\frac{(-)^\lambda}{\lambda}L_{p-\lambda}^{s+\lambda}(x)e_{\lambda-1}(-x) = -\sum_{j=0}^{p-1}\frac{(-x)^j}{j!}\sum_{k=0}^{p-1-j}\frac{1}{k+j+1}\binom{q-1-k}{s+j}$$
$$= \sum_{j=1}^{p}\frac{(-x)^j}{j!}\binom{q}{p-j}\sum_{k=1}^{j}\frac{(k-1)!}{x^k}. \tag{2.16}$$

Here the function $e_m(y)$ is the truncated exponential series

$$e_m(y) = \sum_{i=0}^{m}\frac{y^i}{i!}. \tag{2.17}$$

Now inserting this result into Eq. (2.14) gives us an expression for the radial RCGF when $l \leqslant n-1$ that is in terms of simple functions:

$$g_{nl}(r,r') = \frac{4Z}{n}\frac{p!}{q!}(xx')^l e^{-(x+x')/2}\left\{L_p^s(x)L_p^s(x')\right.$$
$$\times\left[\log x + \log x' + \frac{x+x'}{2n} - \psi(p+1) - \psi(q+1) - \frac{2s+3}{2n} - Ei(x_<)\right]$$
$$+ L_p^s(x)\left[\frac{x'}{n}L_{p-1}^{s+1}(x') + \sum_{k=0}^{p-1}A_k(x')^k - \sum_{k=1}^{s}\binom{q}{p+k}\frac{(k-1)!}{(x')^k}\right]$$
$$+ L_p^s(x')\left[\frac{x}{n}L_{p-1}^{s+1}(x) + \sum_{k=0}^{p-1}A_k x^k - \sum_{k=1}^{s}\binom{q}{p+k}\frac{(k-1)!}{x^k}\right]$$
$$\left. + L_p^s(x_>)e^x\left[\sum_{k=1}^{s}\binom{q-k}{p}\frac{(k-1)!}{x^k} + \Phi_{nl}(x_<)\right]\right\}. \tag{2.18}$$

The Green's functions of Eq. (2.13) satisfy the equation[11]

$$[H_l - E_n]g_{nl}(r,r') = -\frac{\delta(r-r')}{rr'}, \tag{2.19}$$

while those of Eq. (2.18) satisfy

$$[H_l - E_n]g_{nl}(r,r') = R_{nl}(r')R_{nl}(r) - \frac{\delta(r-r')}{rr'}. \tag{2.20}$$

It is convenient here and later to define R_{nl} by Eq. (2.3) for $l \leqslant n-1$ and $R_{nl} = 0$ for $l \geqslant n$. Then both forms of g_{nl} satisfy Eq. (2.20).[29]

B. Integral moments of $g_{nl}(r,r')$

Within (degenerate) RSPT, one encounters integrals of the form

$$\int_0^\infty dr'(r')^2 g_{nl}(r,r')f(r')R_{nl'}(r').$$

In certain select cases,[12-15] the Sturmian representation of g_{nl} can be used to good advantage in evaluating these integrals. In other cases it is easier to use Eqs. (2.13) and (2.18). For the situation in which f is a combination of powers or inverse powers of r', the above integral consists of a linear combination of the integrals

$$I_{nl}^k(x) = \int_0^\infty dx' \mathcal{G}_{nl}(x,x')(x')^{k+2}e^{-x'/2} = \left(\frac{2Z}{n}\right)^{k+3}\int_0^\infty dr' g_{nl}(r,r')(r')^{k+2}e^{-2Zr'/n}, \tag{2.21}$$

where

$$\mathcal{G}_{nl}(x,x') = g_{nl}(r,r'). \tag{2.22}$$

We prefer here to work with the scaled variables $x = 2Zr/n$ and $x' = 2Zr'/n$ for notational simplicity.

It should be mentioned that, for $l \geqslant n$, the indefinite integrals implied in Eq. (2.21) have already been considered by Buchholz.[30]

As pointed out in Ref. 11, a good deal of algebra is involved here; the details of integration over the functions in Eqs. (2.13) and (2.18) and of concomitant reduction to more compact form are fairly lengthy and are given elsewhere.[31] As in the evaluation of radial matrix elements, we find that the integral $I_{nl}^k(x)$ is convergent for $k \geqslant -l-2$, although it must be treated as a principle value integral for certain ranges of k. The results break into six classes as follows:

Case 1: $l \geqslant n$, $k \geqslant l-1$:

$$I_{nl}^k(x) = -\frac{4Z}{n}(l-n)!(l+n)!(k+1-l)!\binom{k+l+2}{l+n}x^{-l-1}e^{-x/2}\sum_{i=2l+1}^{k+l+2}\binom{i-l-n-1}{l-n}\frac{x^i}{i!}; \tag{2.23}$$

Case 2: $l \geqslant n$, $n-2 \leqslant k \leqslant l-2$:

$$I_{nl}^k(x) = (-)^{k+n}\frac{4Z}{n}\frac{(l-n)!(l+n)!}{(l-2-k)!}\binom{k+l+2}{l+n}x^{-l-1}e^{-x/2}\left\{\log x\sum_{i=0}^{l+n}\frac{x^i}{i!}\binom{2l-i}{l-n} - e^x Ei(-x)\sum_{i=0}^{l-n}\frac{(-x)^i}{i!}\binom{2l-i}{l+n}\right.$$
$$-\sum_{i=0}^{l+n}\frac{x^i}{i!}\binom{2l-i}{l-n}\psi(i+1) + \sum_{i=1}^{l-n-1}\frac{x^i}{i!}\sum_{j=1}^{i}\frac{(-)^j}{j!}\binom{2l-i}{l-n-j}$$
$$\left.-\sum_{i=l-n}^{l+n}\frac{x^i}{i!}\binom{2l-i}{l-n}[\psi(2l+1-i)-\psi(l+n+1-i)] + \sum_{i=l+n+1}^{k+2+l}\frac{(-x)^i}{i!}(-)^{l-n}\frac{(2l-i)!(i-l-n-1)!}{(l-n)!}\right\}; \tag{2.24}$$

Case 3: $l \geqslant n$, $-l-2 \leqslant k \leqslant n-3$:

$$I_{nl}^k(x) = -\frac{4Z}{n}\frac{(l-n)!(k+l+2)!(n-3-k)!}{(l-2-k)!}x^{-l-1}e^{-x/2}\sum_{i=k+l+3}^{l+n}\frac{x^i}{i!}\binom{2l-i}{l-n}; \tag{2.25}$$

Case 4: $l \leqslant n-1$, $n-2 \leqslant k$:

$$I_{nl}^k(x) = (-)^{n-l-1}\frac{4Z}{n}\frac{(n-l-1)!(k+l+2)!}{(n+l)!}x^l e^{-x/2}\left(\binom{k+1-l}{n-l-1}\left\{L_{n-l-1}^{2l+1}(x)\left[\frac{x}{2n} + \frac{k-2-3l}{2n} + \gamma\right.\right.\right.$$
$$\left.\left.+\psi(k+3-n)\right] + \frac{x}{n}L_{n-l-2}^{2l+2}(x) - \sum_{i=0}^{n-l-2}\frac{(-x)^i}{i!}\binom{n+l}{n-l-1-i}[\gamma+\psi(n-l-i)]$$
$$\left.+(-)^{n-l}\sum_{i=n-l}^{k+1-l}\frac{x^i}{i!}\frac{(i-n+l)!(n+l)!}{(2l+1+i)!}\right\} - \binom{k+1-l}{n-l-2}\frac{k+l+3}{2n}L_{n-l-1}^{2l+1}(x)\Bigg); \tag{2.26}$$

Case 5: $l < n-1$, $l-1 \leqslant k \leqslant n-3$:

$$I_{nl}^k(x) = \frac{4Z}{n}\frac{(n-l-1)!}{(n+l)!}x^l e^{-x/2}\left[(-)^{k+1-l}\frac{(k+l+2)!(n-3-k)!(k+1-l)!}{(n-l-1)!}\sum_{i=0}^{k+1-l}\frac{(-x)^i}{i!}\binom{n+l}{n-l-1-i}\right.$$
$$\left.+(-)^{n-l}\delta_{k,n-3}\frac{(n+l)!}{2n}L_{n-l-1}^{2l+1}(x)\right]; \tag{2.27}$$

Case 6: $l \leqslant n-1,\ -l-2 \leqslant k \leqslant l-2$:

$$I^k_{nl}(x) = \frac{4Z}{n}\frac{(n-l-1)!(k+l+2)!}{(n+l)!}x^l e^{-x/2} \times\left(\binom{n-3-k}{n-l-1}\left\{L^{2l+1}_{n-l-1}(x)\left[\log x + \frac{x}{2n} + \psi(n-2-k) + \frac{k-3l-2}{2n}\right] + \frac{x}{n}L^{2l+2}_{n-l-2}(x) - \sum_{i=1}^{l-2-k}\binom{n+l}{n-l-1+i}\frac{(i-1)!}{x^i} - \sum_{i=0}^{n-l-1}\frac{(-x)^i}{i!}\binom{n+l}{n-l-1-i}[\psi(i+1)+\psi(i+2l+2)]\right\} + \binom{n-4-k}{n-l-2}\frac{k+l+3}{2n}L^{2l+1}_{n-l-1}(x)\right). \quad (2.28)$$

As a double check on Eqs. (2.23)–(2.28), we note that a comparison of Eqs. (2.20) and (2.21) gives

$$\begin{aligned}[H_l - E_n]I^k_{nl} &= \left(\frac{2Z}{n}\right)^{k+3}\int_0^\infty dr'\,[H_l - E_n]g_{nl}(r,r')(r')^{k+2}e^{-Zr'/n} \\ &= \left(\frac{2Z}{n}\right)^{k+3}\int_0^\infty dr'\left[R_{nl}(r)R_{nl}(r') - \frac{\delta(r-r')}{rr'}\right](r')^{k+2}e^{-Zr'/n} \\ &= \left(\frac{2Z}{n}\right)^{k+3}\left[R_{nl}(r)\int_0^\infty dr'\,R_{nl}(r')(r')^{k+2}e^{-Zr'/n} - r^k e^{-Zr/n}\right] \\ &= (-)^{n-l-1}(k+l+2)!\binom{k+1-l}{n-l-1}\left(\frac{(n-l-1)!Z^3}{(n+l)!}\right)^{1/2}\frac{2}{n^2}R_{nl}(r) - \left(\frac{2Z}{n}\right)^3 x^k e^{-x/2}. \end{aligned} \quad (2.29)$$

Using the facts that $R_{nl} = 0$ for cases (1)–(3) and is given by Eq. (2.3) for cases (4)–(6), and using the identity[32]

$$\binom{j}{i} = (-)^i\binom{-j-1+i}{i}, \quad (2.30)$$

we have been able to prove that each of the above cases obeys Eq. (2.29). It has also been possible to check the orthogonality of g_{nl} and R_{nl} for $l \leqslant n-1$ by use of the results of cases (4) and (5). Cases (1), (4), and (5) agree with the independent derivation by Laurenzi,[33] and cases (1) and (4) reduce (up to a multiplicative factor) to the results of Manakov, Preobrazhenskii, and Rapoport[15] and McDowell.[34]

C. Ladder operator recursion relations

For $l \leqslant n-1$, the ladder operators described by Infeld and Hull[9] can generate $R_{n,l\pm1}$ from R_{nl}. We define these (unnormalized) operators in a slightly different way from their work, however. The most convenient definitions for use here are

$$\hat{\mathscr{L}}^+_{nl}(r) = n\left[1 - \frac{l(l+1)}{Zr} + \frac{(l+1)}{Z}\frac{d}{dr}\right], \quad (2.31)$$

$$\hat{\mathscr{L}}^-_{nl}(r) = n\left[1 - \frac{l(l+1)}{Zr} - \frac{l}{Z}\frac{d}{dr}\right], \quad (2.32)$$

in terms of which

$$\hat{\mathscr{L}}^+_{nl}(r)R_{nl}(r) = -\sqrt{n^2-(l+1)^2}\,R_{n,l+1}(r), \quad (2.33)$$

$$\hat{\mathscr{L}}^-_{nl}(r)R_{nl}(r) = -\sqrt{n^2-l^2}\,R_{n,l-1}(r), \quad l \neq 0. \quad (2.34)$$

In Sec. IV it is shown that, expressing the radial RCGF's in Eq. (2.18) in terms of $r_<$ and $r_>$, if $l \neq n-1$, then

$$g_{n,l+1} = [n^2-(l+1)^2]^{-1}\left[\hat{\mathscr{L}}^+_{nl}(r_>)\hat{\mathscr{L}}^+_{nl}(r_<)g_{nl} - \frac{2n^2(l+1)^2}{Z^2}R_{n,l+1}(r_>)R_{n,l+1}(r_<)\right], \quad (2.35)$$

and, if $l \neq 0$ or n, then

$$g_{n,l-1} = (n^2-l^2)^{-1}\left[\hat{\mathscr{L}}^-_{nl}(r_>)\hat{\mathscr{L}}^-_{nl}(r_<)g_{nl} - \frac{2n^2l^2}{Z^2}R_{n,l-1}(r_>)R_{n,l-1}(r_<)\right]. \quad (2.36)$$

Making the abbreviations

$$(nl\,|\,f(r')\,|\,n'l') = \int_0^\infty dr'\,(r')^2 g_{nl}(r,r')f(r')R_{n'l'}(r'), \quad (2.37)$$

$$\langle nl\,|\,f(r')\,|\,n'l'\rangle = \int_0^\infty dr'\,(r')^2 R_{nl}(r')f(r')R_{n'l'}(r'), \quad (2.38)$$

it is fairly easy to show that Eqs. (2.35) and (2.36) lead to recursion relations such as

(a) $l \neq 0 \neq l', l \neq n = n'$: $(nl|(r')^k|nl') = \frac{\hat{\mathscr{L}}^{+}_{n,l-1}(r)}{n^2 - l^2}\Big[-\sqrt{n^2-(l')^2}\,\frac{l}{l'}(n,l-1|(r')^k|n,l'-1\rangle$

$+ \frac{n}{l'}(l'-l)(n,l-1|(r')^k|nl'\rangle + \frac{ln}{Z}(l'-l-k)(n,l-1|(r')^{k-1}|nl'\rangle\Big]$

$+ \frac{2n^2l^2}{Z^2(n^2-l^2)}[r^kR_{nl'}(r) - \langle nl|(r')^k|nl'\rangle R_{nl}(r)]$; (2.39)

and

(b) $l \neq n-1, n = n'$: $(nl|(r')^k|nl') = \frac{\hat{\mathscr{L}}^{-}_{n,l+1}(r)}{n^2-(l+1)^2}\Big[-\frac{l+1}{l'+1}\sqrt{n^2-(l'+1)^2}(n,l+1|(r')^k|n,l'+1\rangle$

$+ \frac{n(l'-l)}{l'+1}(n,l+1|(r')^k|nl'\rangle + \frac{n(l+1)}{Z}(l'-l-k)(n,l+1|(r')^{k-1}|nl'\rangle\Big]$

$+ \frac{2n^2(l+1)^2}{Z^2[n^2-(l+1)^2]}[r^kR_{nl'}(r) - \langle nl|(r')^k|nl'\rangle R_{nl}(r)]$. (2.40)

The situation for $l \geqslant n$ is even simpler: The corresponding equations are merely given by Eqs. (2.31)–(2.40) with $R_{nl} = 0$ wherever it appears.

Other relations of this sort can be developed without any great difficulty. These may simplify calculations in applications where single integrals over g_{nl} are to be evaluated. In addition, these can lead to recursion relations between double integrals, such as are needed in the calculation of second-order energies. We do not pursue this point further, but instead pass on to a nondifferential recursion relation which is even easier to use.

D. Hypervirial recursion relations

In Sec. V we show how a procedure[10] related to the method of hypervirials[35,36] produces very easily a recursion relation between integrals of the form $(nl|(r')^k|n'l'\rangle$ where only the value of k changes. The result is, for $l \leqslant n-1$ and $k \geqslant -l-l'$,

$$a_{k+2}(nl|(r')^{k+2}|n'l'\rangle + \sum_{j=0}^{2} a_{k-j}(nl|(r')^{k-j}|n'l'\rangle + b_kR_{nl}(r) + \hat{C}_k(r)R_{n'l'}(r) = 0, \tag{2.41}$$

where

$$a_{k+2} = \frac{kZ^4}{2}\left[\frac{1}{(n')^2} - \frac{1}{n^2}\right]^2, \tag{2.42}$$

$$a_k = (k+1)\frac{Z^2}{2}\left\{-k(k+2)\left[\frac{1}{n^2} + \frac{1}{(n')^2}\right] + \left[\frac{1}{n^2} - \frac{1}{(n')^2}\right][l(l+1) - l'(l'+1)]\right\}, \tag{2.43}$$

$$a_{k-1} = Zk(k+2)(2k+1), \tag{2.44}$$

$$a_{k-2} = \frac{k+2}{4}[k^2-(l-l')^2][k^2-(l+l'+1)^2], \tag{2.45}$$

$$b_k = Z^2k\left[\frac{1}{(n')^2} - \frac{1}{n^2}\right]\langle nl|(r')^{k+2}|n'l'\rangle + (k+1)[k(k+2) + l'(l'+1) - l(l+1)]\langle nl|(r')^k|n'l'\rangle, \tag{2.46}$$

$$\hat{C}_k(r) = \frac{kZ^2}{2}\left[\frac{1}{n^2} - \frac{1}{(n')^2}\right]r^{k+2} + \frac{(k+2)}{2}[l(l+1) - l'(l'+1) - k(k-1)]r^k + k(k+2)r^{k+1}\frac{\partial}{\partial r}. \tag{2.47}$$

Originally, the method was used by Epstein, Epstein, and Kennedy[10] to obtain the recursion relation between radial matrix elements (see also Refs. 36 and 37):

$$0 = a_{k+2}\langle nl|(r')^{k+2}|n'l'\rangle + \sum_{j=0}^{2} a_{k-j}\langle nl|(r')^{k-j}|n'l'\rangle, \tag{2.48}$$

where the coefficients are the same as in Eqs. (2.42)–(2.45). This similarity in form is shown to arise from the fact that Eq. (2.20) is an inhomogeneous version of the homogeneous equation satisfied by the radial wave functions, but is nonetheless striking since Eq. (2.48) deals with numbers while Eq. (2.41) deals with functions of r.

We are primarily interested in the case $n = n'$, for which we obtain

$$-k(k+1)(k+2)\frac{Z^2}{n^2}(nl|(r')^k|nl'\rangle + Zk(k+2)(2k+1)(nl|(r')^{k-1}|nl'\rangle$$

$$+ \frac{k+2}{4}[k^2-(l-l')^2][k^2-(l+l'+1)^2](nl|(r')^{k-2}|nl'\rangle$$

$$+ (k+1)[k(k+2) + l'(l'+1) - l(l+1)]\langle nl|(r')^k|nl'\rangle$$

$$+ (k+2)\left[\frac{l(l+1) - l'(l'+1) - k(k-1)}{2}r^k + kr^{k+1}\frac{\partial}{\partial r}\right]R_{nl'}(r) = 0. \tag{2.49}$$

Further specializing to the case $l = l'$, canceling an overall factor of $k\,(k+2)$,

$$-(k+1)\frac{Z^2}{n^2}\langle nl|(r')^k|nl\rangle + Z(2k+1)\langle nl|(r')^{k-1}|nl\rangle + \frac{k}{4}[k^2-(2l+1)^2]\langle nl|(r')^{k-2}|nl\rangle + \left[(k+1)\langle nl|(r')^k|nl\rangle - \frac{k-1}{2}r^k + r^{k+1}\frac{\partial}{\partial r}\right]R_{nl}(r) = 0\,. \tag{2.50}$$

Note that, as with calculations of radial matrix elements, these recursion relations simplify for particular values of k.[10,37] For example, setting $k = 0$ in Eqs. (2.49) and (2.50) gives us, respectively,

$$\frac{l(l+1)-l'(l'+1)}{2}\langle nl|(r')^{-2}|nl'\rangle - \langle nl|nl'\rangle R_{nl}(r) + R_{nl'}(r) = 0 \tag{2.51}$$

and[38]

$$Z\left\langle nl\left|\frac{1}{r'}\right|nl\right\rangle + \left(\frac{3}{2} + r\frac{\partial}{\partial r}\right)R_{nl}(r) = 0\,. \tag{2.52}$$

Thus, we have two sets of integrals "for free", i.e., without recourse to any actual integrations over g_{nl}. In Eq. (2.52) we have used the fact that $\langle nl|nl\rangle = 0$. We now have at our disposal the means to recursively calculate $\langle nl|(r')^k|nl\rangle$ for $k = 1,2,3,\ldots$. However, insertion of $k = -1$ into Eq. (2.50) shows us that this recursion relation does not directly connect $\langle nl|(r')^{-1}|nl\rangle$ to integrals with more negative k.

For $l \neq l'$, using Eq. (2.51) as a starting point, Eq. (2.49) allow us to easily evaluate $\langle nl|(r')^k|nl'\rangle$ for $k = -2, -3,\ldots, -|l-l'|-1$, but not further. Also, we do not know enough from these simple considerations to obtain the integrals with $k \geqslant -1$.

Any particular set of integrals needed to start the recursion process, however, can be constructed from the integrals given in Sec. II B. It is important not to confuse the parameter k from that section with k as used here. As a matter of fact, the relation between the integrals here and the I^k_{nl}'s is[39]

$$\langle nl|(r')^k|nl'\rangle = \left(\frac{n}{2Z}\right)^{k+3}\frac{2}{n^2}\sqrt{\frac{(n-l'-1)!Z^3}{(n+l')!}} \times \sum_{i=0}^{n-l'-1}\frac{(-)^i}{i!} \times \binom{n+l'}{n-l'-1-i} I^{k+l'+i}_{nl}\,. \tag{2.53}$$

We will use Eq. (2.53) in Sec. III, where we calculate $\langle nl|(r')^{-2}|nl'\rangle$ and $\langle nl|(r')^{-3}|nl\rangle$. In certain cases, as mentioned before, it may be easier to use the Sturmian representation. Recursion relations such as in Eqs. (2.39) and (2.40) may also be of help.

Although the preceding has been based on the assumption that $l \leqslant n-1$, the same results hold for $l \geqslant n$ by again setting $R_{nl} = 0$. Of course, this reduces Eqs. (2.50) and (2.52) to trivial statements.

There are other features of interest in these recursion relations, which should be explored in subsequent research. There appears to be a great deal of structure in the integrals considered here, very similar to that discussed in Ref. 37 for radial matrix elements, and a more complete investigation of this structure would be useful.

III. THE SPIN–ORBIT PERTURBATION

In this section we modify the RCGF to include spin, allowing the calculation of the first-order wave function and the second-order energy corresponding to the spin–orbit interaction in a hydrogenic ion.

A. Zeroth-order basis and first-order energy (Refs. 40 and 41)

Our Hamiltonian is taken as

$$H = H_0 + H_1\,, \tag{3.1}$$

where

$$H_0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\left[-\frac{1}{2}\nabla^2 - \frac{Z}{r}\right] \tag{3.2}$$

and

$$H_1 = C\frac{\mathbf{l}\cdot\mathbf{s}}{r^3}\,, \tag{3.3}$$

with C a constant.

As our primitive basis, we take

$$\psi_{nlm_lsm_s}(\mathbf{r}) = R_{nl}(r)Y_{lm_l}(\theta,\varphi)|sm_s\rangle\,. \tag{3.4}$$

Here, $R_{nl}(r)$ is as defined in Eq. (2.3), Y_{lm_l} is chosen with the Condon and Shortley phase, and the spin functions are ($s = \frac{1}{2}$)

$$\left|\frac{1}{2}\,\frac{1}{2}\right\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \quad \left|\frac{1}{2}\,-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\1\end{pmatrix}. \tag{3.5}$$

Coupling $\mathbf{l}$ and $\mathbf{s}$, we obtain $\mathbf{J} = \mathbf{l} + \mathbf{s}$ and the correct zeroth-order basis

$$\chi^{(0)}_{nlsjm_j}(\mathbf{r}) = \sum_{\substack{m_l,m_s\\ m_l+m_s=m_j}}\langle lm_lsm_s|lsjm_j\rangle\psi_{nlm_lsm_s}(\mathbf{r}) = R_{nl}(r)\mathscr{Y}_{lsjm_j}(\theta,\varphi)\,, \tag{3.6}$$

where $\langle lm_lsm_s|lsjm_j\rangle$ is a Clebsch–Gordon coefficient and

$$\mathscr{Y}_{lsjm_j}(\theta,\varphi) = \frac{1}{\sqrt{2l+1}}\begin{pmatrix}\sqrt{l\pm m_j+\frac{1}{2}}\,Y_{l,m_j-\frac{1}{2}}\\ \pm\sqrt{l\pm m_j+\frac{1}{2}}\,Y_{l,m_j+\frac{1}{2}}\end{pmatrix}. \tag{3.7}$$

For convenience, we will drop the subscripts on $\mathscr{Y}$ and $\chi^{(0)}$. We will also strictly neglect the situation $l = 0$, for which we anticipate that the spin–orbit contribution to the energy levels will be zero.[41] Using the orthonormality of the $\chi^{(0)}$'s and

the identity

$$\mathbf{l}\cdot\mathbf{s}\mathscr{Y} = \frac{j(j+1) - l(l+1) - s(s+1)}{2}\mathscr{Y}, \tag{3.8}$$

we obtain the well-known result

$$\begin{aligned} E^{(1)} &= \int_0^\infty \chi^{(0)\dagger} H_1 \chi^{(0)} d\mathbf{r} \\ &= C\lambda_{jls}\int_0^\infty dr\, r^2 R_{nl}(r) r^{-3} R_{nl}(r) \\ &= C\lambda_{jls}\left(\frac{2Z}{n}\right)^3 \frac{(2l-1)!}{(2l+2)!}, \quad l\neq 0, \end{aligned} \tag{3.9}$$

with

$$\lambda_{jls} = \frac{j(j+1) - l(l+1) - s(s+1)}{2}. \tag{3.10}$$

B. The spin-dependent RCGF

Laurenzi[42] has shown tht the three-dimensional spin-dependent RCGF, $\mathsf{G}_n(\mathbf{r},\mathbf{r}')$ is quite simply related to the spin-independent RCGF by

$$\mathsf{G}_n(\mathbf{r},\mathbf{r}') = G_n(\mathbf{r},\mathbf{r}')\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.11}$$

Thus, the first-order wave function is given by ($l\neq 0$)

$$\begin{aligned} \chi^{(1)}(\mathbf{r}) &= \int d\mathbf{r}'\mathsf{G}_n(\mathbf{r},\mathbf{r}')H_1\chi^{(0)}(\mathbf{r}) \\ &= C\lambda_{jls}\mathscr{Y}(\theta,\varphi)(nl\,|(r')^{-3}|nl\rangle, \end{aligned} \tag{3.12}$$

where we have used Eq. (2.11).
The second-order energy is then given by

$$\begin{aligned} E^{(2)} &= \int d\mathbf{r}\chi^{(0)\dagger}H_1\chi^{(1)} \\ &= C^2\lambda_{jls}^2\int_0^\infty dr\, r^2 R_{nl}(r) r^{-3}(nl\,|(r')^{-3}|nl\rangle. \end{aligned} \tag{3.13}$$

C. Evaulation of $\chi^{(1)}$ and $E^{(2)}$

The results of Sec. II D allow us to calculate all of the integrals $(nl\,|(r')^k|nl\rangle$ for $k \geqslant -1$, but not for $k < -1$, so we simultaneously calculate the integrals for $k = -2$ and $k = -3$, providing the beginning of the downward recursive process.

From Eq. (2.53), using the notation $p = n - l - 1$, $s = 2l + 1$, and $q = n + l$,

$$(nl\,|(r')^{-2}|nl\rangle = \frac{1}{nZ}\sqrt{\frac{p!Z^3}{q!}}\sum_{k=0}^{p}\frac{(-)^k}{k!}\binom{q}{p-k}I_{nl}^{l-2+k}. \tag{3.14}$$

Referring to Eqs. (2.27) and (2.28), we have

(a) $1\leqslant k\leqslant n-l-1=p$:

$$\begin{aligned} I_{nl}^{l-2+k} = \frac{4Z}{n}\frac{p!}{q!}x^l e^{-x/2}\Bigg[(-)^{k+1}\frac{(k+s-1)!(p-k)!(k-1)!}{p!}\sum_{\mu=0}^{k-1}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu} \\ + (-)^{p+1}\delta_{k,p}\frac{q!}{2n}L_p^s(x)\Bigg], \end{aligned} \tag{3.15}$$

(b) $k=0$:

$$\begin{aligned} I_{nl}^{l-2} = \frac{4Z}{n}\frac{p!(s-1)!}{q!}x^l e^{-x/2}\Bigg\{L_p^s(x)\left[\log x + \frac{x}{2n} + \psi(p+1) - \frac{s+3}{2n}\right] \\ + \frac{x}{n}L_{p-1}^{s+1}(x) - \sum_{\mu=0}^{p}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu}[\psi(\mu+1)+\psi(s+\mu+1)] + (1-\delta_{p,0})\frac{s}{2n}L_p^s(x)\Bigg\}. \end{aligned} \tag{3.16}$$

So, Eq. (3.14) becomes

$$\begin{aligned} (nl\,|(r')^{-2}|nl\rangle &= \frac{1}{nZ}\sqrt{\frac{p!Z^3}{q!}}\left\{\binom{q}{p}[\text{Eq.(3.16)}] + \sum_{k=1}^{p}\frac{(-)^k}{k!}\binom{q}{p-k}[\text{Eq.(3.15)}]\right\} \\ &= \frac{4}{n^2}\frac{p!}{q!}\sqrt{\frac{p!Z^3}{q!}}x^l e^{-x/2}\Bigg((s-1)!\binom{q}{p}\Bigg\{L_p^s(x)\left[\log x + \frac{x}{2n} + \psi(p+1) - \frac{s+3}{2n}\right] \\ &\quad + \frac{x}{n}L_{p-1}^{s+1}(x) - \sum_{\mu=0}^{p}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu}[\psi(\mu+1)+\psi(s+\mu+1)] + (1-\delta_{p,0})\frac{s}{2n}L_p^s(x)\Bigg\} \\ &\quad - \frac{q!}{p!}\sum_{k=1}^{p}\frac{1}{k(k+s)}\sum_{\mu=0}^{k-1}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu} - \frac{q!}{p!}\frac{1}{2n}(1-\delta_{p,0})L_p^s(x)\Bigg). \end{aligned} \tag{3.17}$$

We can simplify the double summation by interchanging the order and expanding in partial fractions:

$$\begin{aligned} \sum_{k=1}^{p}\frac{1}{k(k+s)}\sum_{\mu=0}^{k-1}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu} &= \sum_{\mu=0}^{p-1}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu}\frac{1}{s}\sum_{k=\mu+1}^{p}\left(\frac{1}{k} - \frac{1}{k+s}\right) \\ &= \frac{1}{s}\sum_{\mu=0}^{p-1}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu}[\psi(p+1) - \psi(\mu+1) - \psi(q+1) + \psi(s+\mu+1)]. \end{aligned} \tag{3.18}$$

Thus,[43]

$$\langle nl|(r')^{-2}|nl\rangle = \frac{4}{sn^2}\sqrt{\frac{p!Z^3}{q!}}x^l e^{-x/2} \times\left\{L_p^s(x)\left[\log x + \frac{x}{2n} + \psi(q+1) - \frac{s+3}{2n}\right] + \frac{x}{n}L_{p-1}^{s+1}(x) - 2\sum_{\mu=0}^{p}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu}\psi(s+\mu+1)\right\}. \tag{3.19}$$

Now, for $l\neq 0$, we find from Eq. (2.31)–(2.34) that

$$\frac{1}{r}R_{nl} = \frac{4Z}{(s-1)(s+1)}R_{nl} + \frac{2Z\sqrt{p(q+1)}}{ns(s+1)}R_{n,l+1} + \frac{2Z\sqrt{q(p+1)}}{ns(s-1)}R_{n,l-1}, \tag{3.20}$$

yielding

$$\langle nl|(r')^{-3}|nl\rangle = \frac{4Z}{(s-1)(s+1)}\langle nl|(r')^{-2}|nl\rangle + \frac{2Z\sqrt{p(q+1)}}{ns(s+1)}\langle nl|(r')^{-2}|n,l+1\rangle + \frac{2Z\sqrt{q(p+1)}}{ns(s-1)}\langle nl|(r')^{-2}|n,l-1\rangle. \tag{3.21}$$

Equation (2.51) gives us the last two integrals immediately, so we have ultimately ($l\neq 0$)

$$\begin{aligned}\langle nl|(r')^{-3}|nl\rangle &= \frac{16Z}{n^2(s-1)s(s+1)}\sqrt{\frac{p!Z^3}{q!}}x^l e^{-x/2}\\ &\times\left\{L_p^s(x)\left[\log x + \frac{x}{2n} + \psi(q+1) - \frac{s+3}{2n} + \frac{p(q+1)(s-1)}{2n^2(s+1)} - \frac{q(p+1)(s+1)}{2n^2(s-1)}\right]\right.\\ &\left. + \frac{x}{n}L_{p-1}^{s+1}(x) - 2\sum_{\mu=0}^{p}\frac{(-x)^\mu}{\mu!}\binom{q}{p-\mu}\psi(s+\mu+1) + \frac{(s-1)}{2n(s+1)}xL_{p-1}^{s+2}(x) - \frac{q(p+1)(s+1)}{2n(s-1)}\frac{1}{x}L_{p+1}^{s-2}(x)\right\}.\end{aligned} \tag{3.22}$$

Inserting this into Eq. (3.12) then provides us with the complete first-order wave function. Using Eq. (3.22), we can evaluate the radial integral

$$\begin{aligned}\int_0^\infty dr\, r^2 R_{nl} r^{-3}\langle nl|(r')^{-3}|nl\rangle &= -\frac{64Z^4}{n^5(s-1)^2s^2(s+1)^2}\left[n^2\left(\frac{1}{s-1}+\frac{1}{s}+\frac{1}{s+1}\right)\right.\\ &\left. + \frac{3}{4}(2q+1) + \frac{9}{8}(2pq+q+p+2)\left(\frac{1}{s-2}+\frac{1}{s+2}\right)\right]\\ &= -\frac{Z^4}{n^5 l^2(l+\frac{1}{2})^2(l+1)^2}\left\{\frac{n^2}{2}\left(\frac{1}{l}+\frac{1}{l+\frac{1}{2}}+\frac{1}{l+1}\right)\right.\\ &\left. + \frac{3}{4}(2n+2l+1) + \frac{9}{16}[2n^2-2l(l+1)+1]\left(\frac{1}{l-\frac{1}{2}}+\frac{1}{l+3/2}\right)\right\}.\end{aligned} \tag{3.23}$$

Coupled with Eq. (3.13), we now have $E^{(2)}$ for any state with $l\neq 0$. Although $E^{(3)}$ is obtainable from knowledge of the first-order wave function, it is unimportant for purposes of our example. The only word of caution we wish to make regards the anomalous magnetic moment of the electron. To first order in the energy, the quantum electrodynamic effects can be approximated by taking C to be

$$C = \frac{Z\alpha^2}{4}g, \tag{3.24}$$

where α is Sommerfeld's fine structure constant and $g = 2.0046...$ However, since the difference $g-2$ is already a perturbation correction, it is not meant to simply be used in higher order.[41]

This example has hopefully illustrated the ease with which RSPT can be performed (even with the inclusion of spin) for any of the hydrogenic bound states. As far as we know, the results of this subsection are completely new.

IV. EVALUATION OF ${}_2F_2(1,1;q+2,p+2;x)$

This section is devoted to the proof of Eq. (2.15). As Laurenzi and Flamberg[11] point out, only a few of these functions are known in closed form, so we go into some detail for the general case. This section may be skipped without loss of continuity.

A. The summation procedure

The function ${}_2F_2(1,1;q+2,p+2;x)$ is a specific case of the confluent hypergeometric series[27]

$${}_2F_2(a,b;c,d;x) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{(c)_k(d)_k}\frac{x^k}{k!}, \tag{4.1}$$

where

$$(\nu)_k = \frac{\Gamma(\nu+k)}{\Gamma(\nu)}, (1)_k = k!. \tag{4.2}$$

We can recast the function we wish to evaluate as a multiple integration over the function $f_j(x)$:

$$f_j(x) = \frac{e^x - e_j(x)}{x^{j+1}}, \tag{4.3}$$

where $e_j(x)$ is the truncated exponential series defined in Eq. (2.17). Instead of ${}_2F_2(1,1;q+2,p+2;x)$ itself, we will consider the equivalent evaluation of

$$\begin{aligned}\Omega^q_{p+1} &= \frac{x^{p+1}}{(p+1)!(q+1)}\,{}_2F_2(1,1;q+2,p+2;x)\\ &= q!\sum_{k=0}^{\infty}\frac{k!}{(q+1+k)!}\frac{x^{p+1+k}}{(p+1+k)!}\\ &= q!\int_0^x dx\ldots\int_0^x dx\, f_q(x)\,. \end{aligned}\tag{4.4}$$

$p+1$ integrations

The following definitions are needed:

$$\mathscr{S}^q_j(x) = (-)^j\sum_{k=0}^{q-j}\binom{q-1-k}{j-1}k!f_k(x)\,, \tag{4.5}$$

$$\mathscr{T}^q_j(x) = (-)^{j+1}L^{q+1-j}_{j-1}(x)[Ei(x) - \log x + \psi(q+1)]\,, \tag{4.6}$$

$$\mathscr{U}^q_{j+1}(x) = (-)^{j+1}\left\{\binom{q}{j}[\gamma+\psi(j+1)] + \sum_{k=1}^{j}\frac{1}{k}\binom{q}{j-k}\right\}, \tag{4.7}$$

$$\mathscr{V}^q_{j+1}(x) = (-)^j\sum_{k=1}^{j}\frac{(-x)^k}{kk!}\binom{q}{j-k}, \tag{4.8}$$

$$\mathscr{W}^q_{j+1}(x) = (-)^j\sum_{k=0}^{j-1}\frac{1}{k+1}\binom{q}{j-1-k}h_k(x)\,, \tag{4.9}$$

and

$$h_k(x) = e^x e_k(-x)\,. \tag{4.10}$$

Also, for an arbitrary function $\psi(x)$, we denote multiple integration by a parenthesized left superscript

$${}^{(r)}\psi(x) = \int_0^x dx\ldots\int_0^x dx\,\psi(x)\,,\ r \text{ integrations} \tag{4.11}$$

with ${}^{(0)}\psi(x) = \psi(x)$.

With these definitions, we will prove shortly that

$$\Omega^q_1(x) = \mathscr{S}^q_1(x) + \mathscr{T}^q_1(x)\,, \tag{4.12}$$

and that

$$\begin{aligned}\int_0^x dx[\mathscr{S}^q_j(x) + \mathscr{T}^q_j(x)] &= \mathscr{S}^q_{j+1}(x) + \mathscr{T}^q_{j+1}(x) + \mathscr{U}^q_{j+1}(x)\\ &\quad + \mathscr{V}^q_{j+1}(x) + \mathscr{W}^q_{j+1}(x)\,. \end{aligned}\tag{4.13}$$

Thus, we find the pattern

$$\begin{aligned}\Omega^q_2 &= \int_0^x dx\Omega^q_1\\ &= \mathscr{S}^q_2 + \mathscr{T}^q_2 + {}^{(0)}\mathscr{U}^q_2 + {}^{(0)}\mathscr{V}^q_2 + {}^{(0)}\mathscr{W}^q_2\,, \end{aligned}\tag{4.14}$$

$$\begin{aligned}\Omega^q_3 &= \int_0^x dx\Omega^q_2\\ &= \mathscr{S}^q_3 + \mathscr{T}^q_3 + {}^{(0)}\mathscr{U}^q_3 + {}^{(0)}\mathscr{V}^q_3 + {}^{(0)}\mathscr{W}^q_3 + {}^{(1)}\mathscr{U}^q_2\\ &\quad + {}^{(1)}\mathscr{V}^q_2 + {}^{(1)}\mathscr{W}^q_2, \end{aligned}\tag{4.15}$$

or, in general,

$$\begin{aligned}\Omega^q_{p+1} &= \mathscr{S}^q_{p+1} + \mathscr{T}^q_{p+1}\\ &\quad + \sum_{r=0}^{p-1}({}^{(r)}\mathscr{U}^q_{p+1-r} + {}^{(r)}\mathscr{V}^q_{p+1-r}\\ &\quad + {}^{(r)}\mathscr{W}^q_{p+1-r})\,. \end{aligned}\tag{4.16}$$

We now turn to the proof of Eqs. (4.12) and (4.13), after which we shall simplify Eq. (4.16).

B. Verification of Eqs. (4.12) and (4.13)

Using the well-known identity[44]

$$\int_0^x dx\, f_0(x) = Ei(x) - \log x - \gamma\,, \tag{4.17}$$

and the obvious identity

$$f_{j-1}(x) = x^{-j}\frac{d}{dx}x^{j+1}f_j(x)\,, \tag{4.18}$$

we find by integration by parts, then iterating,

$$\begin{aligned}\int_0^x dx\, f_q(x) &= -\frac{x}{q}f_q(x) + \frac{1}{q}\int_0^x f_{q-1}(x)\\ &= -x\sum_{k=1}^{q}f_k(x)\frac{(k-1)!}{q!} + \frac{1}{q!}[Ei(x) - \log x - \gamma]\\ &= -\sum_{k=1}^{q}\left[f_{k-1}(x) - \frac{1}{k!}\right]\frac{(k-1)!}{q!} + \frac{1}{q!}[Ei(x) - \log x - \gamma]\\ &= -\frac{1}{q!}\sum_{k=0}^{q}k!f_k(x) + \frac{1}{q!}[Ei(x) - \log x + \psi(q+1)]\,. \end{aligned}\tag{4.19}$$

Reference to Eqs. (4.4)–(4.6) then shows that this is equivalent to Eq. (4.12).

Turning to the proof of Eq. (4.13), Eq. (4.5) is integrated once and compared to Eqs. (4.4) and (4.19) to yield

$$\begin{aligned}\int_0^x dx\mathscr{S}^q_j(x) &= (-)^j\sum_{k=0}^{q-j}\binom{q-1-k}{j-1}\Omega^k_1(x)\\ &= (-)^{j+1}\sum_{i=0}^{q-1-j}i!f_i(x)\sum_{k=i+1}^{q-j}\binom{q-1-k}{j-1}\\ &\quad + (-)^j[Ei(x) - \log x - \gamma]\sum_{k=0}^{q-j}\binom{q-1-k}{j-1}\\ &\quad + (-)^j\sum_{k=1}^{q-j}\binom{q-1-k}{j-1}\sum_{i=1}^{k}\frac{1}{i}\,. \end{aligned}\tag{4.20}$$

In the first two terms we replace k by $q-j-k$ and use Eq. (A1) to obtain

$$\sum_{k=i+1}^{q-j}\binom{q-1-k}{j-1} = \binom{q-i-1}{j}\,, \tag{4.21}$$

$$\sum_{k=0}^{q-j}\binom{q-1-k}{j-1} = \binom{q}{j}\,. \tag{4.22}$$

Use of Eqs. (4.21) and (A2) transforms the last term into

$$(-)^j\sum_{k=1}^{q-j}\binom{q-1-k}{j-1}\sum_{i=1}^{k}\frac{1}{i}$$
$$=(-)^j\sum_{i=1}^{q-j}\frac{1}{i}\sum_{k=i}^{q-j}\binom{q-1-k}{j-1}$$
$$=(-)^j\sum_{i=1}^{q-j}\frac{1}{i}\binom{q-i}{j}=\binom{q}{j}[\psi(q+1)-\psi(j+1)]\,. \tag{4.23}$$

Thus,

$$\int_0^x dx\mathscr{S}_j^q(x)=\mathscr{S}_{j+1}^q(x)+(-)^j\binom{q}{j}[Ei(x)-\gamma-\log x+\psi(q+1)-\psi(j+1)]\,. \tag{4.24}$$

Now we integrate Eq. (4.6) with the help of Eqs. (A3)–(A5):

$$\int_0^x dx\mathscr{T}_j^q(x)=(-)^{j+1}\psi(q+1)\left[\binom{q}{j}-L_j^{q-j}(x)\right]+(-)^{j+1}\sum_{k=0}^{j-1}\frac{(-)^k}{k!}\binom{q}{j-1-k}$$
$$\times\left\{\frac{x^{k+1}}{k+1}[Ei(x)-\log x]+\frac{x^{k+1}}{(k+1)^2}+\frac{(-)^k k!}{(k+1)}[1-e_k(-x)e^x]\right\}$$
$$=(-)^j\left[L_j^{q-j}(x)-\binom{q}{j}\right][Ei(x)-\log x+\psi(q+1)]$$
$$+(-)^{j+1}\sum_{k=1}^{j}\frac{1}{k}\binom{q}{j-k}+(-)^j\sum_{k=1}^{j}\frac{(-x)^k}{kk!}\binom{q}{j-k}$$
$$+(-)^j e^x\sum_{k=0}^{j-1}\frac{1}{k+1}\binom{q}{j-1-k}e_k(-x)\,. \tag{4.25}$$

Adding Eqs. (4.24) and (4.25), and noting the definitions in Eqs. (4.7)–(4.9), we arrive at Eq. (4.13).

C. Simplification of Eq. (4.16)

Inserting the explicit form of $f_k(x)$ into Eq. (4.5) for $\mathscr{S}_{p+1}^q$, doing a little series manipulation, and using Eq. (A6),

$$\mathscr{S}_{p+1}^q=(-)^{p+1}\left[e^x\sum_{k=0}^{q-p-1}\binom{q-1-k}{p}\frac{k!}{x^{k+1}}-\sum_{k=0}^{q-p-1}\binom{q}{p+k+1}\frac{k!}{x^{k+1}}\right]$$
$$=(-)^{p+1}\left[e^x\sum_{k=1}^{s}\binom{q-k}{p}\frac{(k-1)!}{x^k}-\sum_{k=1}^{s}\binom{q}{p+k}\frac{(k-1)!}{x^k}\right]\quad(s=q-p). \tag{4.26}$$

The third term in Eq. (4.16) is easily reduced to

$$\sum_{r=0}^{p-1}{}^{(r)}\mathscr{U}_{p+1-r}^q=-\sum_{r=0}^{p-1}(-)^{p-r}\frac{x^r}{r!}\left\{\binom{q}{p-r}[\gamma+\psi(p+1-r)]+\sum_{k=1}^{p-r}\frac{1}{k}\binom{q}{p-r-k}\right\}$$
$$=(-)^{p+1}\left[\sum_{r=0}^{p-1}\frac{(-x)^r}{r!}\binom{q}{p-r}\sum_{k=1}^{p-r}\frac{1}{k}+\sum_{k=1}^{p}\frac{1}{k}L_{p-k}^{s+k}(x)\right]. \tag{4.27}$$

The fourth term, after manipulating some more series, is

$$\sum_{r=0}^{p-1}{}^{(r)}\mathscr{V}_{p+1-r}^q=(-)^p\sum_{r=0}^{p-1}\sum_{k=1}^{p-r}\frac{(-x)^{k+r}}{k(k+r)!}\binom{q}{p-k-r}=(-)^p\sum_{k=1}^{p}\frac{(-x)^k}{k!}\binom{q}{p-k}\sum_{j=1}^{k}\frac{1}{j}\,. \tag{4.28}$$

For the fifth term in Eq. (4.16), we need to know ${}^{(r)}h_k$. This is

$${}^{(r)}h_k(x)=e^x\sum_{j=0}^{k}\binom{k+r-j}{r}\frac{(-x)^j}{j!}-\sum_{j=0}^{r-1}\binom{k+r-j}{k}\frac{x^j}{j!}\,. \tag{4.29}$$

Although this can be proven directly, it is very simple instead to show that Eq (4.29) satisfies

(i) ${}^{(0)}h_k(x)=e^x e_k(-x)\,,$ (4.30)

(ii) $\frac{d}{dx}[{}^{(r)}h_k(x)]={}^{(r-1)}h_k(x)\,,\qquad r\geqslant 1,$ (4.31)

and

(iii) ${}^{(r)}h_k(0)=0\,,\quad r\geqslant 1.$ (4.32)

Thus,

$$\sum_{r=0}^{p-1}{}^{(r)}\mathscr{W}_{p+1-r}^q=\sum_{r=0}^{p-1}(-)^{p-r}\sum_{k=0}^{p-r-1}\frac{1}{k+1}\binom{q}{p-r-1-k}{}^{(r)}h_k(x)$$
$$=\sum_{k=0}^{p-1}\frac{1}{k+1}\sum_{r=0}^{p-1-k}(-)^{p-r}\binom{q}{p-r-1-k}\left[e^x\sum_{j=0}^{k}\binom{k+r-j}{r}\frac{(-x)^j}{j!}-\sum_{j=0}^{r-1}\binom{k+r-j}{k}\frac{x^j}{j!}\right]$$

$$= \sum_{k=0}^{p-1} \frac{(-)^p}{k+1}\left[e^x \sum_{j=0}^{k} \frac{(-x)^j}{j!} \sum_{r=0}^{p-1-k} (-)^r \binom{q}{p-r-1-k}\binom{k+r-j}{r} - \sum_{j=0}^{p-2-k} \sum_{r=j+1}^{p-1-k} \binom{q}{p-r-1-k}\binom{k+r-j}{k}\right]. \tag{4.33}$$

Now, by some changes of variable and utilization of Eq. (A6), this is transformed into

$$\sum_{r=0}^{p-1} {}^{(r)}\mathscr{W}^{q}_{p+1-r} = (-)^p e^x \sum_{j=0}^{p-1} \frac{(-x)^j}{j!} \sum_{k=0}^{p-1-j} \frac{1}{k+j+1}\binom{q-1-k}{s+j} + (-)^p \sum_{k=0}^{p-1} \frac{1}{k+1} \sum_{j=0}^{p-2-k} \frac{(-x)^j}{j!}\left[\binom{q}{p-1-k-j} - \binom{q-1-k}{p-1-k-j}\right]. \tag{4.34}$$

The sum over j in the second term can be extended to $p-1-k$ without any effect. The second term then becomes, after some more manipulation and use of Eq. (A2),

$$(-)^p\left[\sum_{k=0}^{p-1} \frac{1}{k+1} L^{s+1+k}_{p-1-k}(x) - \sum_{j=0}^{p-1} \frac{(-x)^j}{j!}\binom{q}{p-j} \sum_{k=s+j+1}^{q} \frac{1}{k}\right].$$

Turning to the first term of Eq. (4.34), we define

$$\Phi_{nl}(x) = -\sum_{j=0}^{p-1} \frac{(-x)^j}{j!} \sum_{k=0}^{p-1-j} \frac{1}{k+j+1}\binom{q-1-k}{s+j}. \tag{4.35}$$

The motivation for this definition is that, for $l=0$, this function reduces exactly to the function denoted by a script 0 in Hostler's work.[8] By calculations analogous to his, $\Phi_{n,l}$ may be put in other forms as well, each unfortunately involving a double summation:

$$\Phi_{nl}(x) = \sum_{k=1}^{p-1} \frac{(-)^k}{k!} L^{s+k}_{p-k}(x) e_{k-1}(-x) = \sum_{j=1}^{p} \frac{(-x)^j}{j!}\binom{q}{p-j} \sum_{k=1}^{j} \frac{(k-1)!}{x^k}. \tag{4.36}$$

Thus, Eq. (4.34) is reduced to

$$\sum_{r=0}^{p-1} {}^{(r)}\mathscr{W}^{q}_{p+1-r} = (-)^{p+1} e^x \Phi_{nl}(x) + (-)^p \sum_{k=0}^{p-1} \frac{1}{k+1} L^{s+1+k}_{p-1-k}(x) + (-)^{p+1} \sum_{j=0}^{p-1} \frac{(-x)^j}{j!}\binom{q}{p-j} \sum_{k=s+j+1}^{q} \frac{1}{k}. \tag{4.37}$$

Notice that the second term exactly cancels with the second term in Eq. (4.27).

Collecting the preceding results and inserting them into Eq. (4.16),

$$\Omega^{q}_{p+1}(x) = (-)^{p+1}\left\{e^x\left[\Phi_{nl}(x) + \sum_{k=1}^{s}\binom{q-k}{p}\frac{(k-1)!}{x^k}\right] - \sum_{k=1}^{s}\binom{q}{p+k}\frac{(k-1)!}{x^k} - L^{s}_{p}(x)[Ei(x) - \log x + \psi(q+1)] + \sum_{k=0}^{p-1} \frac{(-x)^k}{k!}\binom{q}{p-k} \sum_{j=s+k+1}^{q} \frac{1}{j} + \sum_{k=0}^{p-1} \frac{(-x)^k}{k!}\binom{q}{p-k} \sum_{j=1}^{p-k} \frac{1}{j} - \sum_{k=1}^{p} \frac{(-x)^k}{k!}\binom{q}{p-k} \sum_{j=1}^{k} \frac{1}{j}\right\}. \tag{4.38}$$

After some minor adjustments, noting the definition of Ω^{q}_{p+1} in Eq. (4.4), we finally arrive at Eq. (2.15).

V. RECURSION RELATIONS FOR $g_{nl}(r,r')$

In their article on the factorization method, Infeld and Hull[9] have shown how ladder operators, essentially those given in Eqs. (2.31) and (2.32), connect the quadratically integrable solutions of the radial hydrogenic equation. These first-order differential operators (first found by Hartree[45]) transform R_{nl} into $R_{n,l\pm1}$.

It is possible to "extend" these operators to energies that are not eigenvalues for application to the inhomogeneous equation satisfied by the radial Green's function $g_l(r,r';E)$. The Whittaker functions in Eq. (2.8) satisfy the relations[46]

$$\left[\nu - \frac{2(l+1)^2}{\rho} + 2(l+1)\frac{d}{d\rho}\right] W_{\nu,l+\frac{1}{2}}(\rho) = (\nu - l - 1) W_{\nu,l+3/2}(\rho), \tag{5.1}$$

$$\left(\nu - \frac{2l^2}{\rho} - 2l\frac{d}{d\rho}\right) W_{\nu,l+\frac{1}{2}}(\rho) = (\nu + l) W_{\nu,l-\frac{1}{2}}(\rho), \tag{5.2}$$

$$\left[\nu - \frac{2(l+1)^2}{\rho} + 2(l+1)\frac{d}{d\rho}\right] \mathscr{M}_{\nu,l+\frac{1}{2}}(\rho) = [(l+1)^2 - \nu^2] \mathscr{M}_{\nu,l+3/2}(\rho), \tag{5.3}$$

$$\left(\nu - \frac{2l^2}{\rho} - 2l\frac{d}{d\rho}\right) \mathscr{M}_{\nu,l+\frac{1}{2}}(\rho) = -\mathscr{M}_{\nu,l-\frac{1}{2}}(\rho). \tag{5.4}$$

With these in hand, we define the operators

$$\hat{\mathscr{L}}^{+}_{\nu,l}(r) = \nu\left[1 - \frac{l(l+1)}{Zr} + \frac{(l+1)}{Z}\frac{\partial}{\partial r}\right], \tag{5.5}$$

$$\hat{\mathscr{L}}^{-}_{\nu,l}(r) = \nu\left[1 - \frac{l(l+1)}{Zr} - \frac{l}{Z}\frac{\partial}{\partial r}\right]. \tag{5.6}$$

Using the notation of Eqs. (2.8) and (2.9), Eqs. (5.1)–(5.4) lead to

$$\hat{\mathscr{L}}^{+}_{\nu,l}(r_>)\frac{1}{\rho_>}W_{\nu,l+\frac{1}{2}}(\rho_>) = (\nu - l - 1)\frac{1}{\rho_>} \times W_{\nu,l+3/2}(\rho_>), \tag{5.7}$$

$$\hat{\mathscr{L}}^{-}_{\nu,l}(r_>)\frac{1}{\rho_>}W_{\nu,l+\frac{1}{2}}(\rho_>) = (\nu + l)\frac{1}{\rho_>}W_{\nu,l-\frac{1}{2}}(\rho_>), \tag{5.8}$$

$$\hat{\mathscr{L}}^{+}_{\nu,l}(r_<)\frac{1}{\rho_<}\mathscr{M}_{\nu,l+\frac{1}{2}}(\rho_<) = [(l+1)^2 - \nu^2]\frac{1}{\rho_<} \times \mathscr{M}_{\nu,l+3/2}(\rho_<), \tag{5.9}$$

$$\hat{\mathscr{L}}^{-}_{\nu,l}(r_<)\frac{1}{\rho_<}\mathscr{M}_{\nu,l+\frac{1}{2}}(\rho_<) = -\frac{1}{\rho_<}\mathscr{M}_{\nu,l-3/2}(\rho_<). \tag{5.10}$$

Combining Eqs. (2.8) and (5.7)–(5.10), we immediately obtain

$$g_{l+1}\left(r,r',-\frac{Z^2}{2\nu^2}\right) = [\nu^2 - (l+1)^2]^{-1}\hat{\mathscr{L}}^{+}_{\nu,l}(r_>) \times \hat{\mathscr{L}}^{+}_{\nu,l}(r_<)g_l\left(r,r';-\frac{Z^2}{2\nu^2}\right), \quad \nu \neq l+1, \tag{5.11}$$

$$g_{l-1}\left(r,r',-\frac{Z^2}{2\nu^2}\right) = (\nu^2 - l^2)^{-1}\hat{\mathscr{L}}^{-}_{\nu,l}(r_>) \times \hat{\mathscr{L}}^{-}_{\nu,l}(r_<)g_l\left(r,r';-\frac{Z^2}{2\nu^2}\right), \quad \nu \neq l \neq 0. \tag{5.12}$$

In the following, we will be interested in the passage to the limit $\nu \to n = 1,2,3\ldots$ In this limit, $\rho_< \to x_< = (2Z/n)r_<$, and the Whittaker functions become, for $l \leqslant n-1$, using Eq. (2.3),[47]

$$\mathscr{M}_{n,l+\frac{1}{2}}(x_<) = \frac{(n-l-1)!}{(n+l)!}x_<^{l+1}e^{-x_</2}L^{2l+1}_{n-l-1}(x_<) = n\sqrt{\frac{(n-l-1)!}{Z(n+l)!}}\,r_<R_{nl}(r_<), \tag{5.13}$$

$$W_{n,l+\frac{1}{2}}(x_>) = (-)^{n-l-1}(n-l-1)!x_>^{l+1}e^{-x_>/2}L^{2l+1}_{n-l-1}(x_>) = (-)^{n-l-1}n\sqrt{\frac{(n-l-1)!(n+l)!}{Z}}\,r_>R_{nl}(r_>). \tag{5.14}$$

At the same time, $\hat{\mathscr{L}}^{\pm}_{\nu,l}(r)$ goes smoothly into $\hat{\mathscr{L}}^{\pm}_{n,l}(r)$, defined in Eqs. (2.31) and (2.32).

Inserting, respectively, Eqs. (5.11) and (5.12) into Eq. (2.12), we obtain

$$\begin{aligned} g_{n,l+1}(r,r') &= \frac{n}{2}\left\{\frac{\partial}{\partial(1/\nu)}\left(\frac{1}{\nu^2} - \frac{1}{n^2}\right)[\nu^2 - (l+1)^2]^{-1} \times \hat{\mathscr{L}}^{+}_{\nu,l}(r_>)\hat{\mathscr{L}}^{+}_{\nu,l}(r_<)g_l\left(r,r';-\frac{Z^2}{2\nu^2}\right)\right\}_{\nu=n} \\ &= \frac{\hat{\mathscr{L}}^{+}_{n,l}(r_>)\hat{\mathscr{L}}^{+}_{n,l}(r_<)}{n^2-(l+1)^2}\frac{n}{2}\left[\frac{\partial}{\partial(1/\nu)}\left(\frac{1}{\nu^2} - \frac{1}{n^2}\right)g_l\right]_{\nu=n} \\ &\quad + \frac{\hat{\mathscr{L}}^{+}_{n,l}(r_>)\hat{\mathscr{L}}^{+}_{n,l}(r_<)}{2n}\left\{\frac{2(l+1)^2}{\nu}\left[1-\left(\frac{l+1}{\nu}\right)^2\right]^{-2} \times\left(\frac{1}{\nu^2} - \frac{1}{n^2}\right)g_l\right\}_{\nu=n} \\ &= \frac{\hat{\mathscr{L}}^{+}_{n,l}(r_>)\hat{\mathscr{L}}^{+}_{n,l}(r_<)}{n^2-(l+1)^2}\left\{g_{n,l} + \frac{n^2(l+1)^2}{n^2-(l+1)^2} \times\lim_{\nu\to n}\left[\left(\frac{1}{\nu^2} - \frac{1}{n^2}\right)g_l\right]\right\}, \quad l \neq n-1, \end{aligned} \tag{5.15}$$

and

$$g_{n,l-1}(r,r') = \frac{\hat{\mathscr{L}}^{-}_{n,l}(r_>)\hat{\mathscr{L}}^{-}_{n,l}(r_<)}{n^2-l^2}\left\{g_{n,l} + \frac{n^2l^2}{n^2-l^2} \times\lim_{\nu\to n}\left[\left(\frac{1}{\nu^2} - \frac{1}{n^2}\right)g_l\right]\right\}, \quad n \neq l \neq 0. \tag{5.16}$$

The remaining limit term in the above is simply $(-2/Z^2)$ times the residue of g_l at $E = E_n$, i.e.,

$$\lim_{\nu\to n}\left[\left(\frac{1}{\nu^2} - \frac{1}{n^2}\right)g_l\left(r,r';-\frac{Z^2}{2\nu^2}\right)\right] = -\frac{2}{Z^2}R_{nl}(r)R_{nl}(r'). \tag{5.17}$$

Thus, referring back to Eqs. (2.33) and (2.34),

$$\begin{aligned} g_{n,l+1}(r,r') &= \frac{\hat{\mathscr{L}}^{+}_{n,l}(r_>)\hat{\mathscr{L}}^{+}_{n,l}(r_<)}{n^2-(l+1)^2}\left\{g_{n,l}(r,r') - \frac{2n^2(l+1)^2}{Z^2[n^2-(l+1)^2]}R_{nl}(r_>)R_{nl}(r_<)\right\} \\ &= \frac{\hat{\mathscr{L}}^{+}_{n,l}(r_>)\hat{\mathscr{L}}^{+}_{n,l}(r_<)}{n^2-(l+1)^2}g_{nl}(r,r') - \frac{2n^2(l+1)^2}{Z^2[n^2-(l+1)^2]}R_{n,l+1}(r_>) \times R_{n,l+1}(r_<), \end{aligned} \tag{5.18}$$

$$\begin{aligned} g_{n,l-1}(r,r') &= \frac{\hat{\mathscr{L}}^{-}_{n,l}(r_>)\hat{\mathscr{L}}^{-}_{n,l}(r_<)}{n^2-l^2}\left[g_{nl}(r,r') - \frac{2n^2l^2}{Z^2(n^2-l^2)}R_{nl}(r_>)R_{nl}(r_<)\right] \\ &= \frac{\hat{\mathscr{L}}^{-}_{n,l}(r_>)\hat{\mathscr{L}}^{-}_{n,l}(r_<)}{n^2-l^2}g_{nl}(r,r') - \frac{2n^2l^2}{Z^2(n^2-l^2)}R_{n,l-1}(r_>)R_{n,l-1}(r_<). \end{aligned} \tag{5.19}$$

For purposes of these operations, g_{nl} is considered as a function of $r_<$ and $r_>$. Following Hostler,[8] we write $r_>$ and $r_<$ in terms of the Heaviside theta function

$$\theta(y) = \begin{cases} 1, & y > 0, \\ 0, & y < 0. \end{cases} \tag{5.20}$$

$$d\theta(y)/dy = \delta(y), \tag{5.21}$$

$$r_> = \theta(r-r')r + \theta(r'-r)r', \tag{5.22}$$

$$r_< = \theta(r-r')r' + \theta(r'-r)r, \tag{5.23}$$

$$\partial r_> /\partial r = \theta(r-r'), \quad \partial r_< /\partial r = \theta(r'-r). \tag{5.24}$$

Then, Eq. (2.20) separates into the three relations

$r > r'$:

$$\left[-\frac{1}{2r_>}\frac{\partial^2}{\partial r_>^2} r_> + \frac{l(l+1)}{2r_>^2} - \frac{Z}{r_>} + \frac{Z^2}{2n^2}\right] g_{nl} = R_{nl}(r_<)R_{nl}(r_>); \tag{5.25}$$

$r < r'$:

$$\left[-\frac{1}{2r_<}\frac{\partial^2}{\partial r_<^2} r_< + \frac{l(l+1)}{2r_<^2} - \frac{Z}{r_<} + \frac{Z^2}{2n^2}\right] g_{nl} = R_{nl}(r_<)R_{nl}(r_>); \tag{5.26}$$

$r = r'$:

$$\left[\left(\frac{\partial}{\partial r_>} - \frac{\partial}{\partial r_<}\right) g_{nl}\right]_{r=r'} = \frac{2}{r^2}. \tag{5.27}$$

Formally, use of Eqs. (5.22) – (5.24) also relates the ladder operators as

$$\hat{\mathscr{L}}^{\pm}_{n,l}(r_>)\hat{\mathscr{L}}^{\pm}_{n,l}(r_<) = \theta(r-r')\hat{\mathscr{L}}^{\pm}_{nl}(r)\hat{\mathscr{L}}^{\pm}_{nl}(r') + \theta(r'-r)\hat{\mathscr{L}}^{\pm}_{nl}(r')\hat{\mathscr{L}}^{\pm}_{nl}(r). \tag{5.28}$$

Now, assuming $f(r')$ to be a continuous function, we define the integral $(0 \neq l \neq n)$

$$\mathscr{I} = \int_0^\infty dr'(r')^2 f(r') g_{nl}(r,r'). \tag{5.29}$$

By Eq. (5.18) this is

$$\begin{aligned}
\mathscr{I} &= \int_0^\infty dr'(r')^2 f(r') \left[\frac{\hat{\mathscr{L}}^{+}_{n,l-1}(r_>)\hat{\mathscr{L}}^{+}_{n,l-1}(r_<)}{n^2-l^2} g_{n,l-1}(r,r') - \frac{2n^2l^2}{Z^2(n^2-l^2)} R_{nl}(r)R_{nl}(r')\right] \\
&= (n^2-l^2)^{-1}\int_0^\infty dr'(r')^2 f(r')[\theta(r-r')\hat{\mathscr{L}}^{+}_{n,l-1}(r)\hat{\mathscr{L}}^{+}_{n,l-1}(r') + \theta(r'-r)\hat{\mathscr{L}}^{+}_{n,l-1}(r')\hat{\mathscr{L}}^{+}_{n,l-1}(r)] g_{n,l-1}(r,r') \\
&\quad - \frac{2n^2l^2}{Z^2(n^2-l^2)}\int_0^\infty dr'(r')^2 f(r') R_{nl}(r')R_{nl}(r) \\
&= (n^2-l^2)^{-1}\left[\int_0^r dr'(r')^2 f(r')\hat{\mathscr{L}}^{+}_{n,l-1}(r)\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r')\right. \\
&\quad \left. + \int_r^\infty dr'(r')^2 f(r')\hat{\mathscr{L}}^{+}_{n,l-1}(r')\hat{\mathscr{L}}^{+}_{n,l-1}(r) g_{n,l-1}(r,r') - 2n^2l^2Z^{-2}R_{nl}(r)\int_0^\infty dr'(r')^2 f(r')R_{nl}(r')\right].
\end{aligned} \tag{5.30}$$

Pulling $\hat{\mathscr{L}}^{+}_{n,l-1}(r)$ out of the integrals, taking note of the integration limits,

$$\begin{aligned}
\mathscr{I} = (n^2-l^2)^{-1}\Bigg[&\hat{\mathscr{L}}^{+}_{n,l-1}(r)\int_0^\infty dr'(r')^2 f(r')\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r') - \frac{nl}{Z}\lim_{\substack{r'\to r\\ r'<r}} f(r')(r')^2\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r') \\
&+ \frac{nl}{Z}\lim_{\substack{r'\to r\\ r'>r}} f(r')(r')^2\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r') - 2\frac{n^2l^2}{Z^2}R_{nl}(r)\int_0^\infty dr'(r')^2 f(r')R_{nl}(r')\Bigg].
\end{aligned} \tag{5.31}$$

Since g_{nl} is continuous at $r = r'$, the limit terms are only nonzero because of the factor $(nl/Z)(\partial/\partial r')$ in $\hat{\mathscr{L}}^{+}_{n,l-1}(r')$; so

$$\begin{aligned}
&-\frac{n}{Z}\lim_{\substack{r'\to r\\ r'<r}} f(r')(r')^2\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r') + \frac{n}{Z}\lim_{\substack{r'\to r\\ r'>r}} f(r')(r')^2\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r') \\
&= \frac{n^2l^2}{Z^2} f(r) r^2\left[\left(\frac{\partial}{\partial r_>} - \frac{\partial}{\partial r_<}\right) g_{n,l-1}\right]_{r=r'} = \frac{2n^2l^2}{Z^2} f(r)
\end{aligned} \tag{5.32}$$

by Eq. (5.27).

Thus, for $0 \neq l \neq n$,

$$\begin{aligned}
\mathscr{I} &= \int_0^\infty dr'(r')^2 f(r') g_{nl}(r,r') \\
&= (n^2-l^2)^{-1}\hat{\mathscr{L}}^{+}_{n,l-1}(r)\int_0^\infty dr'(r')^2 f(r')\hat{\mathscr{L}}^{+}_{n,l-1}(r') g_{n,l-1}(r,r') \\
&\quad + \frac{2n^2l^2}{Z^2(n^2-l^2)}\left[f(r) - R_{nl}(r)\int_0^\infty dr'(r')^2 f(r')R_{nl}(r')\right].
\end{aligned} \tag{5.33}$$

Proceeding along similar lines we find that for $l \neq n-1$,

$$\begin{aligned}
\mathscr{I} &= [n^2-(l+1)^2]^{-1}\hat{\mathscr{L}}^{-}_{n,l+1}(r)\int_0^\infty dr'(r')^2 f(r')\hat{\mathscr{L}}^{-}_{n,l+1}(r') g_{n,l+1}(r,r') \\
&\quad + \frac{2n^2(l+1)^2}{Z^2[n^2-(l+1)^2]}\left[f(r) - R_{nl}(r)\int_0^\infty dr'(r')^2 f(r')R_{nl}(r')\right].
\end{aligned} \tag{5.34}$$

Equations (2.39) and (2.40) are obtained from Eqs. (5.33) and (5.34) very easily by setting $f(r') = (r')^k R_{nl'}(r')$, integrating by parts, and using simple identities such as

$$\hat{\mathscr{L}}_{n,l}(r') = \frac{nl}{l'}\left[\frac{l'}{l} - \frac{l'(l+1)}{Zr'} - \frac{l'}{Z}\frac{\partial}{\partial r'}\right] = \frac{l}{l'}\hat{\mathscr{L}}_{n,l'}(r') + \frac{n(l'-l)}{l'} + \frac{nl(l'-l)}{Zr'}. \tag{5.35}$$

For $l \geqslant n$, it is unnecessary to repeat the derivations of this section: Excluding Eqs. (5.13) and (5.14), all of the results of this section carry over by setting $R_{nl} = 0$. For example, Eq. (5.17) is then merely the statement that $g_l(r,r';E)$ does not have a pole at $E = E_n$, a fact already discussed in Sec. II.A.

VI. HYPERVIRIAL RECURSION RELATIONS FOR $\langle nl|(r')^k|n'l'\rangle$

Following arguments parallel to those in Ref. 10, we consider integrals involving $g_{nl}(r,r')$, $R_{n'l'}(r')$, and the quantity $H_l W - WH_{l'}$. H_l is defined in Eq. (2.7) and $W(r', p_{r'})$ is a function of r' and the radial "momentum" operator $p_{r'} = -(i/r')(\partial/\partial r')r'$.

Expressing H_l and $H_{l'}$ in terms of the variable r', we have that

$$[H_{l'} - E_{n'}]R_{n'l'}(r') = 0, \tag{6.1}$$

$$[H_l - E_n]g_{nl}(r,r') = R_{nl}(r)R_{nl}(r') - \frac{\delta(r-r')}{rr'}. \tag{2.20}$$

Since n, n', l and l' do not change in the following, we abbreviate the notation as

$$(f(r', p_{r'})) = (nl\,|f(r', p_{r'})|\,n'l'), \tag{6.2}$$

$$\langle f(r', p_{r'})\rangle = \langle nl\,|f(r', p_{r'})|\,n'l'\rangle, \tag{6.3}$$

where the right-hand sides of these are defined in Eqs. (2.37) and (2.38), respectively.

Now, for a reasonable function W, we can utilize Eqs. (6.1) and (2.20) and integration by parts to get

$$(H_l W - WH_{l'}) = [E_n - E_{n'}](W) + \langle W\rangle R_{nl}(r) - W(r, p_r)R_{n'l'}(r). \tag{6.4}$$

For the particular choice $W(r') = (r')^k$,

$$H_l W - WH_{l'} = -ikp_{r'}(r')^{k-1} + \frac{k(k-1) + l(l+1) - l'(l'+1)}{2}(r')^{k-2}. \tag{6.5}$$

Equating the two forms of $(H_l W - WH_{l'})$ arising from Eqs. (6.4) and (6.5), we get

$$-ik(p_{r'}(r')^{k-1}) = [E_n - E_{n'}]((r')^k) - \frac{k(k-1) + l(l+1) - l'(l'+1)}{2}((r')^{k-2}) + \langle (r')^k\rangle R_{nl}(r) - r^k R_{n'l'}(r). \tag{6.6}$$

For the second choice $W(r', p_{r'}) = p_{r'}(r')^{k+1}$,

$$H_l W - WH_{l'} = -i(k+1)p_{r'}^2(r')^k + \frac{k(k+1) + l(l+1) - l'(l'+1)}{2}p_{r'}(r')^{k-1} - il(l+1)(r')^{k-2} + iZ(r')^{k-1}. \tag{6.7}$$

Further, in the integration of Eq. (6.7) we can integrate by parts for the term in $p_{r'}^2$, and then use Eq. (2.20) and the fact that

$$p_{r'}^2 = 2\left[H_l - \frac{l(l+1)}{2(r')^2} + \frac{Z}{r'}\right] \tag{6.8}$$

to obtain

$$\begin{aligned}(H_l W - WH_{l'}) = {} & -2iE_n(k+1)((r')^k) + i(k+1)l(l+1)((r')^{k-2}) - 2iZ(k+1)((r')^{k-1}) - 2i(k+1)\langle (r')^k\rangle R_{nl}(r) \\ & + 2i(k+1)r^k R_{n'l'}(r) + \frac{k(k+1) + l(l+1) - l'(l'+1)}{2}((p_{r'}r')^{k-1}) \\ & - il(l+1)((r')^{k-2}) + iZ((r')^{k-1}).\end{aligned} \tag{6.9}$$

This choice of W is then inserted into Eq. (6.4), and $(H_l W - WH_{l'})$ eliminated between Eqs. (6.4) and (6.9), to give a recursion relation involving $p_{r'}$ and powers of r'. Elimination of integrals with derivatives sandwiched between g_{nl} and $R_{n'l'}$ is accomplished by Eq. (6.6).[48] The result of the uninteresting algebra is

$$\begin{aligned}0 = {} & k(E_n - E_{n'})^2((r')^{k+2}) + \{k(k+1)(k+2)(E_n + E_{n'}) + (k+1)(E_{n'} - E_n)[l(l+1) - l'(l'+1)]\}((r')^k) \\ & + Zk(k+2)(2k+1)((r')^{k-1}) + \frac{k+2}{4}\{-4k^2 l(l+1) \\ & + [k(k+1) + l(l+1) - l'(l'+1)][k(k-1) + l(l+1) - l'(l'+1)]\}((r')^{k-2}) \\ & + \Big\{k(E_n - E_{n'})\langle (r')^{k+2}\rangle - ik(k+2)\langle p_{r'}(r')^{k+1}\rangle\end{aligned}$$

$$+(k+2)\left[2k(k+1)-\frac{k(k+1)+l(l+1)-l'(l'+1)}{2}\right]\Bigg\}R_{nl}(r)$$
$$+\left\{k(E_{n'}-E_n)r^{k+2}+ik(k+2)p_r r^{k+1}+(k+2)\left[\frac{k(k+1)+l(l+1)-l'(l'+1)}{2}-2k(k+1)\right]r^k\right\}R_{n'l'}(r). \tag{6.10}$$

From Ref. 10, the matrix element analog of Eq. (6.6),

$$-ik\langle p_r(r')^{k-1}\rangle=(E_n-E_{n'})\langle(r')^k\rangle-\frac{k(k-1)+l(l+1)-l'(l'+1)}{2}\langle(r')^{k-2}\rangle, \tag{6.11}$$

permits a simplification of the coefficient of $R_{nl}(r)$. After a little more algebraic manipulation, we end up with Eq. (2.41). Although we have not done so, it is also possible to simplify radial derivatives of $R_{n'l'}(r)$ by use of Eqs. (2.33) and (2.34).

ACKNOWLEDGMENTS

The authors wish to thank Professor Richard A. Askey and Professor John E. Harriman for very helpful conversations, and Professor Bernard J. Laurenzi for sending us a preprint of Ref. 33 after the completion of our manuscript. This research was supported by National Science Foundation Grant CHE77-18754 and one of us (BRJ) is also very grateful to the American Oil Company for an AMOCO fellowship.

APPENDIX A

In this Appendix, identities and integrals are listed as needed:

$$\sum_{k=0}^{n}\binom{k+m}{k}=\binom{n+m+1}{m+1}\quad\text{(Ref. 49)}, \tag{A1}$$

$$\sum_{k=1}^{q-j}\frac{1}{k}\binom{q-k}{j}=\binom{q}{j}[\psi(q+1)-\psi(j+1)]\quad\text{(Ref. 50)}, \tag{A2}$$

$$\int_0^x dx\,x^k Ei(x)=\frac{x^{k+1}}{k+1}Ei(x)+\frac{(-)^k k!}{k+1}[1-e^x e_k(-x)],\quad k\geqslant 0, \tag{A3}$$

$$\int_0^x dx x^k\log x=\frac{x^{k+1}}{k+1}\log x-\frac{x^{k+1}}{(k+1)^2},\quad k\geqslant 0, \tag{A4}$$

$$\int_0^x dx L_{j-1}^{q+1-j}(x)=\binom{q}{j}-L_j^{q-j}(x)\quad\text{(Ref. 51)}, \tag{A5}$$

$$\sum_{j=0}^{n}(-)^j\binom{\alpha+n}{n-j}\binom{j+k}{j}=\binom{\alpha+n-1-k}{n}\quad\text{(Ref. 52)}. \tag{A6}$$

[Equations (A3) and (A4) easily follow by integration by parts.]

APPENDIX B. DIRECT EVALUATION OF $g_{n,n-1}(r,r')$

In this Appendix, we use the results of Sec. V to show how the function $g_{nl}(r,r')$ can be obtained in closed form for $l\leqslant n-1$, once $g_l(r,r';E)$ has been obtained. This treatment is similar in spirit ot the method used by White[53] to quickly evaluate $G_1(\mathbf{r},\mathbf{r}';E)$. For the radial wave functions R_{nl}, $l=n-1$ is considered the top of the ladder in l, with

$$\hat{\mathscr{L}}^{+}_{n,n-1}(r)R_{n,n-1}(r)=0. \tag{B1}$$

For $l=n-1$, Eq. (4.11) becomes

$$g_n\left(r,r';-\frac{Z^2}{2\nu^2}\right)$$
$$(\nu^2-n^2)^{-1}\hat{\mathscr{L}}^{+}_{\nu,n-1}(r_>)\hat{\mathscr{L}}^{+}_{\nu,n-1}(r_<)g_{n-1}\left(r,r';-\frac{Z^2}{2\nu^2}\right). \tag{B2}$$

Multiplying both sides by (ν^2-n^2) and inserting both sides into the derivative operation in Eq. (2.12),

$$\frac{n}{2}\left[\frac{\partial}{\partial(1/\nu)}(\nu^2-n^2)\left(\frac{1}{\nu^2}-\frac{1}{n^2}\right)g_n\right]_{\nu=n}$$
$$=\frac{n}{2}\left[\frac{\partial}{\partial(1/\nu)}\left(\frac{1}{\nu^2}-\frac{1}{n^2}\right)\hat{\mathscr{L}}^{+}_{\nu,n-1}(r_>)\right.$$
$$\left.\times\hat{\mathscr{L}}^{+}_{\nu,n-1}(r_<)g_{n-1}\right]_{\nu=n}. \tag{B3}$$

Since $g_n(r,r';-Z^2/2\nu^2)$ does not have a pole at $\nu=n$, the left-hand side is zero:

$$0=\frac{\hat{\mathscr{L}}^{+}_{n,n-1}(r_>)\hat{\mathscr{L}}^{+}_{n,n-1}(r_<)}{2n}\left[\frac{\partial}{\partial(1/\nu)}\left(\frac{1}{\nu^2}-\frac{1}{n^2}\right)\nu^2 g_{n-1}\right]_{\nu=n}$$
$$=\frac{\hat{\mathscr{L}}^{+}_{n,n-1}(r_>)\hat{\mathscr{L}}^{+}_{n,n-1}(r_<)}{2n}\left\{n^2\left[\frac{\partial}{\partial(1/\nu)}\left(\frac{1}{\nu^2}-\frac{1}{n^2}\right)g_{n-1}\right]_{\nu=n}-2\left[\left(\frac{1}{\nu^2}-\frac{1}{n^2}\right)\nu^3 g_{n-1}\right]_{\nu=n}\right\}$$
$$=\hat{\mathscr{L}}^{+}_{n,n-1}(r_>)\hat{\mathscr{L}}^{+}_{n,n-1}(r_<)\left[g_{n,n-1}(r,r')+\frac{2n^2}{Z^2}R_{n,n-1}(r_<)R_{n,n-1}(r_>)\right]$$
$$=\hat{\mathscr{L}}^{+}_{n,n-1}(r_>)\hat{\mathscr{L}}^{+}_{n,n-1}(r_>)g_{n,n-1}(r,r') \tag{B4}$$

by Eq. (B1).

For simplicity, we now switch to the variables $x_<$ and $x_>$. Recalling that $\mathscr{G}_{n,n-1}(x_<,x_>)=g_{n,n-1}(r_<,r_>)$, we assume that

$$\mathscr{G}_{n,n-1}(x_<,x_>)=(x_<x_>)^{n-1}e^{-(x_<+x_>)/2}f(x_<,x_>), \tag{B5}$$

so that Eq. (B4) becomes

$$0=\left[n-\frac{2n(n-1)}{x_>}+2n\frac{\partial}{\partial x_>}\right]\left[n-\frac{2n(n-1)}{x_<}+2n\frac{\partial}{\partial x_<}\right]\mathcal{G}_{n,n-1}(x_<,x_>)$$
$$=(x_<x_>)^{n-1}e^{-(x_<+x_>)/2}\frac{\partial}{\partial x_>}\frac{\partial}{\partial x_<}f(x_<,x_>). \quad \text{(B6)}$$

This implies that $f(x_<,x_>)$ is sum separable:

$$f(x_<,x_>)=\tau(x_>)+\Psi(x_<). \quad \text{(B7)}$$

Equations (5.25) – (5.27) give us three equations for f:

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial x_>^2}+\left(\frac{1}{2}-\frac{n}{x_>}\right)\frac{\partial}{\partial x_>}\right]f(x_<,x_>)=\frac{Z}{n^2(2n-1)!}, \quad \text{(B8)}$$

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial x_<^2}+\left(\frac{1}{2}-\frac{n}{x_<}\right)\frac{\partial}{\partial x_<}\right]f(x_<,x_>)=\frac{Z}{n^2(2n-1)!}, \quad \text{(B9)}$$

$$\left[\left(\frac{\partial}{\partial x_>}-\frac{\partial}{\partial x_<}\right)f(x_<,x_>)\right]_{x=x'}=\frac{4Z}{n}x^{-2n}e^x. \quad \text{(B10)}$$

We now shift all of the asymmetry into $\Psi(x_<)$, i.e.,

$$\Psi(x_<)=\tau(x_<)+\Lambda(x_<). \quad \text{(B11)}$$

Then Eqs. (B8) – (B10) tell us that

$$\left[-\frac{1}{2}\frac{d^2}{dx^2}+\left(\frac{1}{2}-\frac{n}{x}\right)\frac{d}{dx}\right]\tau(x)=\frac{Z}{n^2(2n-1)!}, \quad \text{(B12)}$$

$$\left[-\frac{1}{2}\frac{d^2}{dx^2}+\left(\frac{1}{2}-\frac{n}{x}\right)\frac{d}{dx}\right]\Lambda(x)=0, \quad \text{(B13)}$$

$$\frac{\partial}{\partial x}\Lambda(x)=-\frac{4Z}{n}x^{-2n}e^x. \quad \text{(B14)}$$

Equation (B14) may be integrated immediately to obtain

$$\Lambda(x)=\frac{4Z}{n(2n-1)!}\left[e^x\sum_{k=1}^{2n-1}\frac{(k-1)!}{x^k}-Ei(x)\right]+\text{const.}, \quad \text{(B15)}$$

which satisfies both Eqs. (B13) and (B14). Now let

$$\chi(x)=\frac{d}{dx}\tau(x). \quad \text{(B16)}$$

Equation (B12) becomes

$$\left[-\frac{1}{2}\frac{d}{dx}+\left(\frac{1}{2}-\frac{n}{x}\right)\right]\chi(x)=\frac{Z}{n^2(2n-1)!}. \quad \text{(B17)}$$

Then letting

$$\chi(x)=x^{-2n}e^x\Delta(x), \quad \text{(B18)}$$

we get

$$\frac{d}{dx}\Delta(x)=-\frac{2Z}{n^2(2n-1)!}x^{2n}e^{-x}, \quad \text{(B19)}$$

or (A = constant)

$$\Delta(x)=\frac{4Z}{n}e_{2n}(x)e^{-x}+A. \quad \text{(B20)}$$

Thus,

$$\chi(x)=\frac{4Z}{n}x^{-2n}e_{2n}(x)+Ax^{-2n}e^x, \quad \text{(B21)}$$

$$\tau(x)=\int^x\chi(y)dy+\text{const.} \quad \text{(B22)}$$

From Eq. (B21) we see that $\tau(x_>)$ will blow up as $x_>\to\infty$ unless $A=0$. Thus,

$$\tau(x)=\frac{4Z}{n(2n-1)!}\left[\frac{x}{2n}+\log x-\sum_{k=1}^{2n-1}\frac{(2n-1)!x^{-k}}{k(2n-1-k)!}\right]+\text{const.} \quad \text{(B23)}$$

Compiling the results of Eqs. (B5), (B7), (B11), (B15), and (B23), we obtain

$$g_{n,n-1}(r_<,r_>)=\mathcal{G}_{n,n-1}(x_<,x_>)=\frac{4Z}{n(2n-1)!}(x_<x_>)^{n-1}e^{-(x_<+x_>)/2}\left[\frac{x_<+x_>}{2n}+\log x_<+\log x_>-\sum_{k=1}^{2n-1}\frac{(2n-1)!}{k(2n-1-k)!}(x_<^{-k}+x_>^{-k})+\text{const.}+e^x\sum_{k=1}^{2n-1}\frac{(k-1)!}{x_<^k}-Ei(x_>)\right]. \quad \text{(B24)}$$

The constant term is then uniquely determined by the orthogonality condition between $g_{nl}(r,r')$ and $R_{nl}(r')$, turning out to be $\gamma-2-\psi(2n+1)$.

[1]H.F. Hameka, J. Chem. Phys. **47**, 2728 (1967); **48**, 4810 (1968).
[2]B.J. Laurenzi, Chem. Phys. Lett. **1**, 641 (1968).
[3]L. Hostler, Phys. Rev. **178**, 126 (1969).
[4]A.I. Sherstyuk, Opt. Spectrosc. **30**, 191 (1971).
[5]A.I. Sherstyuk, Theor. Math. Phys. **7**, 579 (1971).
[6]L. Swierkowski and M. Suffczynski, Bull. Acad. Polon. Sci. **21**, 285 (1973).
[7]S.J. Niemczyk, Chem. Phys. Lett. **20**, 146 (1973).
[8]L. Hostler, J. Math. Phys. **16**, 1585 (1975).
[9]L. Infeld and T.E. Hull, Rev. Mod. Phys. **23**, 21 (1951).
[10]S.T. Epstein, J.H. Epstein, and B. Kennedy, J. Math. Phys. **8**, 1747 (1966).
[11]B.J. Laurenzi and A. Flamberg, Int. J. Quantum Chem. **11**, 869 (1977).
[12]S.V. Khristenko and S.I. Vetchinkin, Opt. Spectrosc. **31**, 269 (1971).
[13]A.I. Sherstyuk, Theor. Math. Phys. **21**, 1097 (1974).
[14]A.F. Shestakov, Opt. Spectrosc. **44**, 357 (1978).
[15]N.L. Manakov, M.A. Prebrazhenskii, and L.P. Rapoport, Opt. Spectrosc. **35**, 14 (1973).

[16]Atomic units are used: $\hbar = \mu = e = 1, \mu$ = reduced mass.
[17]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), 4th edition.
[18]H. Bethe and E. Salpeter, *Quantum Mechanics of One- and Two-Electron Systems* (Academic, New York, 1957), Sec. 4.
[19]R.A. Mapleton, J. Math. Phys. **2**, 178 (1961). The symbol $S_{n'}$, as usual, indicates summation over the discrete radial wave functions and integration over the continuum ones. For notational compactness, it is easier not to explicitly introduce a new symbol for the continuum functions.
[20]L. Hostler, J. Math. Phys. **5**, 591 (1964). The notation and normalization is slightly different in the present work. Note also the ν as defined by Eq. (2.10) is doublevalued for $E > 0$, but this is not important to the considerations in this paper.
[21]H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969). The Whittaker functions defined here differ slightly from those, say, in Ref. 17.
[22]See J.O. Hirschfelder, W.B. Brown, and S.T. Epstein, Adv. Quantum Chem. **1**, 257 (1964) and references therin.
[23]A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, p. 136.
[24]Throughout this work, we adhere to the convention that a sum reduces to zero when the upper index is less than the lower one. Also, log$x$ is intended to represent the natural logarithm of x.
[25]Note that our definitions of A_k and B_k differ by the factor of $q!$ from those of Ref. 4.
[26]Reference 23, Vol. 1, Chap. 1.
[27]Reference 17, p. 1045.
[28]Refernece 17, Sec. 8.2.
[29]Multiples of the Whittaker functions provide solutions of the radial eigenvalue equation if r is not allowed to span the entire range $0 \leqslant r < \infty$, and there is then no reason to stipulate that n is a positive integer greater than l. Since we are not considering a restricted range here, there are no radial wave functions that are quadratically integrable for $l \geqslant n$, but many of the equations we present for $l \leqslant n-1$ hold for $l \geqslant n$ if we use the artifice of setting $R_{nl} = 0$.
[30]Refernece 21, pp. 37–39, 117; also see H. Buchholz, Math. Z. **57**, 167 (1953). I^k_{nl} here is directly related to Buchholz' $R^{k+5/2}_{n,l+1/2}$.
[31]B.R. Johnson, University of Wisconsin Theoretical Chemistry Institute Reports WIS-TCI-615. This report is solely devoted to the details involved in deriving Eqs. (2.23)–(2.28).
[32]H.W. Gould, *Combinatorial Identities* (Morgantown, Morgantown, 1972), p. iv.
[33]B.J. Laurenzi, Int. J. Quantum. Chem. **15**, 717 (1979).
[34]H.K. McDowell, Chem. Phys. Lett. **52**, 145 (1977).
[35]J.O. Hirschfelder, J. Chem. Phys. **33**, 1762 (1960); S.T. Epstein and J.O. Hirschfelder, Phys. Rev. **123**, 1495 (1961); J.H. Epstein and S.T. Epstein, Am. J. Phys. **30**, 266 (1962); J.C.Y. Chen, J. Chem. Phys. **40**, 615 (1964).
[36]D.E. Hughes, J. Phys. B **10**, 3167 (1977).
[37]P. Blanchard, J. Phys. B **7**, 993 (1974).
[38]Other means of evaluating these integrals exists. In particular, Eq. (2.52) is already known from the general fact that a perturbation λ/r for a hydrogenic ion is equivalent to a dilation of the nuclear charge from Z to $Z - \lambda$. The first-order wave function essentially corresponds to differentiation with respect to Z of the zeroth-order wave function. See P.D. Robinson, J. Chem. Phys. **47**, 2319 (1967); and J. Killingbeck, Rep. Prog. Phys. **40**, 963 (1977).
[39]A close scrutiny of Eqs. (2.39), (2.40), and (2.53) might lead one to expect certain of the integrals $\langle nl|(r')^k|nl'\rangle$ to contain terms such as $\log(2Zr/n)$ for l and $l' \leqslant n-1, l \neq l'$, and $k = -2, -3, \ldots, -|l-l'|-1$. This seems to contradict the statement that we can generate these integrals directly from R_{nl}, $R_{nl'}$, and Eq. (2.51) since there are no logarithmic terms involved. However, it is fairly easy to show that the logarithmic term is multiplied by a coefficient proportional to $\langle nl|(r')^k|nl'\rangle$; the Pasternak–Sternheimer identity [S. Pasternak and R.M. Steinheimer, J. Math. Phys. **3**, 1280 (1962)] states that $\langle nl|(r')^k|nl'\rangle = 0$ for $l \neq l'$, $k = -2, -3, \ldots, -|l-l'|-1$. Thus, no logarithmic terms appear until $k = -|l-l'|-2$, and this can be seen to be consistent with the recursion relation (2.49). Similar breaks in the chain can be found for $l \geqslant n$.
[40]Reference 18, Sec. 13.
[41]J.E. Harriman, *Theoretical Foundations of Electron Spin Resonance* (Academic, New York, 1978), Chap. 1, Sec. 2 and 3.
[42]B.J. Laurenzi, J. Chem. Phys. **56**, 619 (1972).
[43]The integral in Eq. (3.19) corresponds to a modification of $l(l+1)$ in the radial Hamiltonian. Within the radial subspace, this corresponds to a dilation of l while keeping $n - l$ constant (the generalized Kepler problem of Ref. 10), much like the dilation of Z mentioned in Ref. 38. By expressing l and n in R_{nl} [Eq. (2.3)] in terms of $p = n - l - 1$ and $s = 2l + 1$, we can recover Eq. (3.19) through differentiation with respect to s.
[44]Reference 17, p. 927.
[45]D.R. Hartree, Proc. Cambridge Philos. Soc. **25**, 310 (1929).
[46]Reference 21, p. 82.
[47]Reference 21, pp. 13 and 23.
[48]Much the same as in Ref. 7 of Ref. 10, we can generalize the results here to the case where W also contains a factor $e^{-\beta r}$, leading to recurrences between the integrals like $\langle (r')^k e^{-\beta r}\rangle$ and $\langle p_r (r')^k e^{-\beta r}\rangle$. However, it does not appear possible to eliminate the second type of integral between the resulting relations except in the case $n = n'$, $l = l'$.
[49]Reference 17, p. 3.
[50]Reference 32, p. 17.
[51]Reference 17, p. 843.
[52]Reference 32, p. 22.
[53]R.J. White, Phys. Rev. **184**, 245 (1969).

Does there exist a scattering theory for time automorphism groups of C^*-algebras corresponding to two-body interactions?[a)]

H. Narnhofer
Institut für Theoretische Physik, Universität Wien, Austria

(Received 13 June 1979; accepted for publication 22 August 1979)

It is shown that at least on the level of perturbation theory there does not exist a scattering automorphism between free time evolution and time evolution corresponding to two-body interaction for the C^*-algebra of fermions on a lattice system.

1. INTRODUCTION

In a few particle systems scattering theory is a useful tool to describe the behavior of particles since it tells us that the motion can be transformed into the free motion and especially the number of constants of the motion is unchanged (at least in the relevant channel). In Ref. 1 the method of scattering theory was first transposed to automorphism groups on C^* algebras. Especially, the idea of a intertwining automorphism was introduced, combining two automorphism groups by

$$\tau_t = \gamma \tau_t^0 \gamma^{-1}.$$

For instance, the existence of

$$\lim_{t \to \pm \infty} \tau_t \tau^0_{-t} A = \gamma_{\pm} A,$$

$$\lim_{t \to \pm \infty} \tau^0_{-t} \tau_t A = \gamma_{\pm}^{-1} A, \quad \forall A \in \mathcal{A},$$

would lead to such an intertwining automorphism. Such an automorphism would imply the following: (a) If τ_t^0 is asymptotically Abelian, so is τ_t. (b) If ω^0 is τ_t^0 (extremal) invariant, then $\omega^0 \circ \gamma^{-1}$ is τ_t (extremal) invariant. (c) If ω^0 is τ_t^0 − KMS, then $\omega \circ \gamma^{-1}$ is τ_t − KMS. (d) If σ_z^0 is an automorphism group that commutes with τ_t^0, then $\gamma \sigma_z^0 \gamma^{-1}$ commutes with τ_t. This last property can be understood as corresponding to the statement in few particle physics on the constants of motion.[2,3]

So it occurs that the existence of such an intertwining automorphism is in contradiction to the hope of the physicist that ergodic properties are improved by the interaction of particles. In this note we do not pretend to show that such an intertwining automorphism between free time evolution and time evolution corresponding to two body interactions does not exist. However, we will show that at least it cannot be obtained on the level of perturbation theory because we are led to divergences. That perturbations of fourth power in the field operators are already drastic is nothing new. In fact, already in this order the condition of dynamical stability implied the KMS condition for free particles.[4,5] Here local perturbations were considered whereas we take only global perturbations into account. Another example where divergences occur in perturbation theory, already for the six-point functions (in dimension 1 and 2, otherwise higher order has to be taken into account) is the attempt to generalize the Boltzmann equation for higher densities.[6] Our approach differs insofar as Ref. 6 treats the classical problem and considers states and the limit $t \to \infty$, whereas we stay on the algebraic level and use the stationary approach.

2. THE MODEL AND THE RESULT

In order that everything is well defined we restrict ourselves on the Fermi lattice system, e.g., we consider the C^* algebra built by creation and annihilation operators a_x^+, and a_y, respectively, satisfying

$$[a_x^+, a_y]_+ = \delta_{xy},$$

$$[a_x, a_y]_+ = 0, \quad x, y \text{ in a } r \text{ dimensional lattice}.$$

The free time evolution reads

$$\tau_t^0 a(f) = a(e^{iht} f)$$

and corresponds to a derivation

$$\delta_0 [a(f)] = \sum_{x,y} h(x-y)[a_x^+ a_y, a(f)]_-,$$

whereas the time evolution τ_t^λ with interaction results from the derivation

$$\begin{aligned}(\delta_0 + \lambda \delta_v)[a(f)] &= \sum_{x,y} h(x-y)[a_x^+ a_x^+, a(f)]_- \\ &+ \sum_{x_1,x_2,y_1,y_2} \lambda\, v(x_1 - x_2, x_2 - y_1, y_1 - y_2) \\ &\times [a_{x_1}^+ a_{x_2}^+ a_{y_1} a_{y_2}, a(f)]_- \\ &= [H_0, a(f)]_- + \lambda [V, a(f)]_-,\end{aligned}$$

such that τ_t commutes with space translations. v has to have decreasing properties, such that the domain of δ_v contains the strictly local operators (δ_V is approximately inner).[6] Due to the anticommutation relations and the self-adjointness of V we can assume

$$\begin{aligned}v(x_1 - x_2, x_2 - y_1, y_1 - y_2) &= -v(x_1 - x_2, x_1 - y_2, y_2 - y_1) \\ &= v(x_2 - x_1, x_1 - y_2, y_2 - y_1), \text{ etc.}\end{aligned}$$

or for the Fourier transformation

$$\begin{aligned}\tilde{v}(p_1, p_2, p_3) &= \tilde{v}(p_3, p_2, p_1)^* = -\tilde{v}(p_1, p_2, p_2 - p_3) \\ &= -\tilde{v}(p_2 - p_1, p_2, p_3).\end{aligned}$$

[a)]Work supported in part by "Fonds zur Förderung der wissenschaftlichen Forschung in Österreich," Project nr. 3569.

 0022-2488/79/122502-04$01.00

Our result reads as follows:

Theorem: There does not exist a family intertwining automorphisms γ_λ of A, such that (a) for $\lambda < \lambda_0$,

$$\tau_t^\lambda = \gamma_\lambda \tau_t^0 \gamma^{-1}$$

or equivalently

$$\delta_0 + \lambda \delta_V = \gamma_\lambda \delta_0 \gamma_\lambda^{-1}.$$

(b) γ_λ is continuous in the sense that

$$\lim_{\lambda \to 0} \| \gamma_\lambda A - A \| = 0, \quad \forall A.$$

Under these assumptions we can write $\gamma_\lambda = e^{\lambda \epsilon(\lambda)}$, where $\epsilon(\lambda)$ is a λ dependent derivation, and can be written as $\epsilon_1 + 0(1)$. (c) ϵ_1 satisfies: $\cup_\Lambda A_\Lambda \subset \mathcal{D}(\epsilon_1)$.

Remark: In two body systems Moeller operators are known to satisfy the corresponding continuity conditions (in fact, they are usually analytic in λ), where the domain on which differentiation is possible is given as the domain of H_0, even if scattering theory in the strict sense does not work, but one has to consider modified Moeller operators. Therefore, the above assumptions seem reasonable.

The chosen domain for the derivation guarantees that for given $A \in \mathcal{A}_\Lambda$, $\epsilon_1(A)$ can be written as [Ref. 7, Eqs. (3.2.52)]

$$\epsilon_1(A) = \epsilon_1^{(n)}(A) = i[E_1^{(n)}, A],$$

with $A \in \mathcal{A}_\Lambda$, $E_1^{(n)} \in \mathcal{A}_{\Lambda_n}$, $\Lambda_n \supset \Lambda$ and sufficiently large. Since τ_t and δ_V are invariant under space translations, also ϵ_1 had to be invariant. It has to satisfy

$$\delta_V = [\epsilon_1, \delta_0].$$

It follows that ϵ_1 corresponds to a (formal) operator

$$E_1 = \sum_{x_1, x_2, y_1, y_2} g(x_1 - x_2, x_2 - y_1, y_1 - y_2) a_{x_1}^+ a_{x_2}^+ a_{y_1} a_{y_2}$$

which satisfies

$$i[E_1, H_0] = V.$$

Due to the anticommutation relations and self-adjointness, g is restricted by

$$\begin{aligned} \tilde{g}(p_1, p_2, p_3) &= -\tilde{g}(p_2 - p_1, p_2, p_3) \\ &= -\tilde{g}(p_1, p_2, p_2 - p_3) = \tilde{g}(p_3, p_2, p_1)^*. \end{aligned}$$

A straightforward calculation leads to the equality

$$\tilde{g}(p_1, p_2, p_3) = \frac{i\tilde{v}(p_1, p_2, p_3)}{\tilde{h}(p_1) + \tilde{h}(p_2 - p_1) - \tilde{h}(p_3) - \tilde{h}(p_2 - p_3)}.$$

This expression corresponds to the first term in expansion in powers of λ of the Moeller operator

$$\begin{aligned} \Omega' &= \lim_{t \to \infty} i \int_0^t e^{iH_0 t'} V e^{-iH_0 t'} dt' \\ &= \lim_{\epsilon \downarrow 0} \int dE \frac{1}{H_0 - E + i\epsilon} V \delta(H_0 - E) \end{aligned}$$

and this already tells us that we have to wonder whether we can make the singularity in $\tilde{g}$ meaningful, namely, with the help of a δ-type distribution. We will show that this is impossible. Take the observable a_z^+. Then

$$\begin{aligned} &[E_1, a_z^+] \\ &= 2 \int dp_1 dp_2 dp_3 \tilde{g}(p_1, p_2, p_3) a_{p_1}^+ a_{p_2 - p_1}^+ a_{p_2 - p_3}. \end{aligned}$$

To show that this is unbounded we estimate $\| \Sigma f(x_1, x_2, y_1) a_{x_1}^+ a_{x_2}^+ a_{y_1} \|$. To do this we consider the expectation value in the trace state. That we have to turn to another representation than the Fock representation is not surprising since the lack of scattering theory should be caused by the existence of infinitely many particles. This trace state, being the quasifree state with two point function

$$\omega(a_x^+ a_y) = \tfrac{1}{2} \delta_{xy},$$

can be considered as Fock state over a C^* algebra, constructed by creation and annihilation operators α_x^+, β_y^+, α_x, β_y satisfying[8]

$$\begin{aligned} \alpha_x &= 2^{-1/2}[\pi(a_x) - \pi(b_x^+)], \\ \pi(a_x) &= 2^{-1/2}(\alpha_x + \beta_x^+), \\ \beta_y &= 2^{-1/2}[\pi(a_y^+) + \pi(b_y)], \\ \pi(b_x^+) &= 2^{-1/2}(-\alpha_x + \beta_x^+), \end{aligned}$$

where $\pi(b_x^+)\pi(b_y)$ belongs to the commutant $\pi(\mathcal{A})'$.

Therefore,

$$\begin{aligned} \pi(a_{x_1}^+ a_{x_2}^+ a_{y_1}) = 2^{-3/2} a_{x_1}^+ a_{x_2}^+ \beta_{y_1}^+ \\ + \text{terms containing annihilation operators.} \end{aligned}$$

Taking the expectation value with the α–β vacuum and assuming $f(x_1, x_2, y_1) = -f(x_2, x_1, y_1)$, we obtain

$$\| \sum f(x_1, x_2, y_1) a_{x_1}^+ a_{x_2}^+ a_{y_1} \| \geqslant 2^{-1} \| f \|_2.$$

The denominator of $\tilde{g}$ vanishes on a hyperplane in the $3r$ dimensional space; therefore, g does not belong to L^2 and the $\epsilon_1 a_z^+$ is not a quasilocal operator.

The same ideas can be used to state the next theorem.

Theorem: Let $\bar{\delta}_0$ be a derivation implementing an automorphism group that commutes with δ_0. Then there does not exist a family of derivations $\bar{\delta}(\lambda)$ satisfying the following:

(a) $[\bar{\delta}(\lambda), \delta_0 + \lambda \delta_V] = 0$;

(b) $\bar{\delta}(\lambda)$ is differentiable on $\cup_\Lambda \mathcal{A}_\Lambda$.

Evidently, if we had found a family γ_λ, this would define such a family, namely, $\bar{\delta}_\lambda = \gamma_\lambda \bar{\delta}_0 \gamma_\lambda^{-1}$. On the other hand, this result is stronger. It is interesting also from another point of view, considering the τ_t^0 invariant states: If such a state ω_0 is cyclic and separating, it is KMS with respect to some τ_t^0. If we could perturb ω_0 in such a way that it stays cyclic and separating and the corresponding KMS automorphism would be an automorphism group not only of the von Neumann algebra but of the C^* algebra itself, we would have already a family of $\bar{\delta}_\lambda$. Therefore, our result is a strong hint that perturbation theory is only possible for KMS states and not for general invariant states. (For local perturbation this is the equivalence of dynamical stability and the KMS condition which is already shown under appropriate assumptions.[5,7,9]

The proof is the same as before: We know that $\overline{\delta_0}$ is quasifree; therefore, the considered commutators are of the same type. Writing

$$\bar{\delta}(\lambda) = \overline{\delta_0} + \lambda\delta_1 + 0(\lambda),$$

we have

$$[\overline{\delta_0}, \delta_V] = [\delta_0, \delta_1]$$

such that δ_1 corresponds to a $\bar{g}$

$$\bar{\tilde{g}}(p_1,p_2,p_3) = \{\tilde{v}(p_1,p_2,p_3)[\bar{\tilde{h}}(p_1) + \bar{\tilde{h}}(p_2 - p_1) - \bar{\tilde{h}}(p_2 - p_3) - \bar{\tilde{h}}(p_3)]\}/[\tilde{h}(p_1) + \tilde{h}(p_2 - p_1) - \tilde{h}(p_2 - p_3) - \tilde{h}(p_3)].$$

Again the denominator vanishes on a hyperplane that does not coincide with the hyperplane of zeros of the nominator. Therefore, by the same argument as for ϵ_1, δ_1 is not well defined.

3. SCATTERING THEORY IN REPRESENTATIONS

The question arises whether scattering theory might not work in a weaker sense. Instead of considering norm convergence we might consider strong convergence in representations. To make this meaningful we have to consider representations where both τ_t^0 and τ_t^λ are unitarily implemented. This is not the case in general. The exceptions follow.

A. The Fock representation

Since both τ_t^0 and τ_t^λ commute with the gauge transformation, i.e., the number of particles does not change, the question whether scattering theory exists or not for the von Neumann algebra, i.e., if

$$\operatorname{st}\lim_{t \to \pm\infty} e^{iHt}e^{-iH_0t} = \Omega_\pm,$$

$$\operatorname{st}\lim_{t \to \pm\infty} e^{iH_0t}e^{-iHt} = \Omega^+,$$

exists, can be solved on the n-particle level. This is already a hard and in general unsolved problem in few particle physics. However, for continuous systems with usual free time evolution and two body interactions, decreasing sufficiently fast for large distances, it was shown in Ref. 10 that generically scattering theory exists for n particles, n arbitrary. Therefore, the existence of an intertwining automorphism on the weak closure of the C^* algebra in the Fock representation seems to be a good conjecture, at least for small perturbations.

It seems worthwhile to consider what the difference is between the representation and results on the C^* algebra: Accordingly, our E_1 should be a properly defined, though unbounded operator in this representation. In fact,

$$\lim_{\epsilon \downarrow 0} \int dp_1\,dp_2\,dp_3 \times \frac{\tilde{v}(p_1\,p_2\,p_3)}{\tilde{h}(p_1) + \tilde{h}(p_2 - p_1) - \tilde{h}(p_2 - p_3) - \tilde{h}(p_3) \pm i\epsilon} \times a^+_{p_1}a^+_{p_2 - p_1}a_{p_2 - p_3}a_{p_3}\prod a^+(f_i)|\Omega\rangle$$

is a well defined vector of finite norm for a dense set of f_i, because the singularity is integrated over before $|\,|^2$ is performed. Also, higher derivations remain harmless. Consider, for example, ϵ_2 defined by

$$[\epsilon_2, \delta_0] = -\tfrac{1}{2}[\epsilon_1, [\epsilon_1, \delta_0]].$$

The part corresponding to three creation and annihilation operators defined by

$$[E_{26}, H_0] = \tfrac{1}{2}:[V, E_1]$$

corresponds to a

$$\tilde{d}(p_1,p_2,p_4 - p_3 + p_2, p_4, p_5) = \{2[\tilde{v}(p_1,p_2,p_3)\tilde{g}(p_3,p_4,p_5) - \tilde{g}(p_1,p_2,p_3) \times \tilde{v}(p_3,p_4,p_5)]\}[\tilde{h}(p_1) + \tilde{h}(p_2 - p_1) + \tilde{h}(p_4 - p_3) - \tilde{h}(p_2 - p_3) - \tilde{h}(p_4 - p_5) - \tilde{h}(p_5)]^{-1}.$$

Evidently, the singularity now seems worse, but its action on a vector corresponds to two convolutions and again the singularity becomes harmless.

B. The trace state

The state is obtained as $\omega(A) = \lim_{\Lambda \to \infty} \operatorname{Tr}_{\mathcal{H}_\Lambda} A / \operatorname{Tr}_{\mathcal{H}_\Lambda} 1$. Therefore, it can be considered as an equilibrium state at infinite temperature for any approximately inner time evolution. The corresponding von Neumann algebra is of type II_1.

The free time evolution is formally implemented by

$$H_0 = \sum f(x - y)\pi(a_x)^+\pi(a_y) - \sum f(x - y)\pi(b_x)^+\pi(b_y).$$

Expressed by the creation and annihilation operators α and β, respectively, introduced in Sec. 2, this leads to

$$H_0 = \sum f(x - y)(\alpha_x^+\alpha_y - \beta_x^+\beta_y).$$

In the same way we obtain

$$V = \sum v(x_1,x_2,y_1,y_2)(\alpha^+_{x_1}\alpha^+_{x_2}\alpha_{y_1}\beta^+_{y_2} + \text{corresponding terms}),$$

where only terms with an odd number of creation and annihilation operators occur. Since now the perturbation is of an essentially different structure than for the Fock representation, we cannot adopt the results of Ref. 10 to conclude that scattering theory works. However, at least we can consider

$$\operatorname{st}\lim_{t \to +\infty} e^{iHt}e^{-iH_0t}\prod\alpha^+(f_i)\beta^+(g_j)|\Omega\rangle = \operatorname{st}\lim_{t \to +\infty} i\int_0^t dt'\,e^{iHt'}Ve^{-iH_0t'} \times \prod\alpha^+(f_i)\beta^+(g_j)|\Omega\rangle - \prod\alpha^+(f_i)\beta^+(g_j)|\Omega\rangle.$$

In the Fock vacuum V always contained a term with two annihilation operators, so that by the coupling of V the weak

convergence of $e^{iht} f_i$ and $e^{iht} g_j$ was sufficient to imply integrability of the relevent vector. Now we have also to consider a term with three creation and only one annihilation operator, and here the relevant term is

$$\sum_{y_1} v(x_1 - x_2, x_2 - y_2, y_2 - y_1)(e^{iht} f)(y_1),$$

which does not converge strongly to zero with $t \to \infty$.

[1]D.W Robinson, Commun. Math. Phys. **31**, 171 (1973).
[2]H. Narnhofer, Acta Phys. Austr. **47**, 1 (1977).
[3]W. Thirring, *Classical Dynamical Systems* (Springer, New York, 1978).
[4]E. Trych Pohlmeier, Thesis (1974).
[5]R. Haag and E. Trych Pohlmeier, Commun. Math. Phys. **56**, 213 (1977).
[6]E.G.D. Cohen in *The Physicist's Conception of Nature*, edited by J. Mehra (Reidel, Dordrecht, 1973), p. 548.
[7]O. Bratteli and D.W. Robinson "C^*- and W^*-Algebras and Applications to Quantum Statistical Mechanics," in *Texts and Monographs in Physics* (Springer, Heidelberg, 1979).
[8]R. Haag, D. Kastler, and E. Trych Pohlmeier, Commun. Math. Phys. **38**, 173 (1974).
[9]H. Narnhofer, "Scattering Theory for Quasifree Time Automorphisms of C-Algebras and von Neumann Algebras," Rep. Math. Phys. (to be published).
[10]I. Sigal, Mem. Am. Math. Soc. **16**, 209 (1978).

Positive and negative frequency decompositions in curved spacetime

Prakash Panangaden
Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 8 May 1979; accepted for publication 25 July 1979)

In this note we derive a formula for the positive and negative frequency parts of a solution in terms of the Feynman propagator. Our arguments are valid in the presence of particle creation. We also derive a formula for an operator $\mathscr{F}$, that gives the particle creation rate. The formalism uses complex structures to capture the notion of positive and negative frequencies and thus avoids using analyticity arguments. The results obtained clarify the relation between approaches to quantum field theory based on the complex structure and approaches in which the propagator is the basic object. We will consider only scalar fields for simplicity.

I. INTRODUCTION

In this note we derive a formula for the complex structures associated with a quantum field theory in curved space–time. The action of the complex structures as well as the action of an operator describing particle creation are given in terms of the Feynman propagator. In an approach to field theory begun by Segal[1] and by Lichnerowicz,[2] the complex structure is used to capture the notion of positive and negative frequency decompositions of solutions of the wave equation. This approach has been extended by Ashtekar and Magnon-Ashtekar[3,4] and by Kay[5] to include particle creation effects. Unfortunately, their constructions are difficult to carry out in practice. The main results of this note is to give concrete expressions to their abstractly defined operators. These formulas allow one to see the relation between the approach to field theory of Segal[1] and Ashtekar and Magnon-Ashtekar[3,4] based on complex structures and the approach of Schwinger,[6] De Witt,[7] and Rumpf[8] based on Feynman propagators.

An application of these results would be to check that a propagator does define a positive and negative frequency decomposition. It has become popular to use Euclideanization[9,10] to define a propagator. In this approach the definitions of positive and negative frequencies are not explicit and it is of interest to obtain them. This is especially true in cosmological space–times where the definition of "early time" particle states is physically obscure.[10]

II. DEFINITIONS

Let M be a globally hyperbolic Lorentzian manifold of class C^∞ with a C^∞ metric g_{ab} defined on it. The free, neutral scalar field of mass m is described by the Klein–Gordon equation

$$(\Box - m^2)\phi = 0, \tag{1}$$

where $\Box$ is the Laplace–Beltrami operator and the field $\phi(x)$ is a real function on M. The Cauchy problem for Eq. (1) and for data on a space like hypersurface Σ is solved by[2]

$$\phi(x) = \int_\Sigma \{\phi(y)\,\nabla^a_y D(x,y) - D(x,y)\nabla^a\phi(y)\}d\sigma_a(y), \tag{2}$$

where $d\sigma^a$ is the volume element on Σ and $D(x,y)$ is the difference between the advanced and retarded Green functions of Eq. (1). $D(x,y)$ is skewed in its arguments. Associated with solutions of Eq. (1) is a canonical, skewed two-form Ω called the symplectic structure:

$$\Omega[\phi_1(x),\phi_2(x)] = \int_\Sigma \{\phi_1\,\nabla^a\phi_2 - \phi_2\,\nabla^a\phi_1\}d\sigma_a. \tag{3}$$

Because ϕ_1 and ϕ_2 are solutions of Eq. (1), Ω is independent of the hypersurface Σ.

The Feynman propagator $G_F(x,x')$ is defined as a (distributional) solution to

$$(\Box - m^2)G_F(x,x') = -\delta(x,x'), \tag{4}$$

which is symmetric in x and x'. To obtain a unique solution to Eq. (4) we must of course impose boundary conditions. If we have a positive and negative frequency decomposition (e.g., static space–times), then we can impose the "causal" boundary condition; positive frequencies are propagated into the future while negative frequencies are propagated into the past. If there is no such canonical decomposition, then one is forced to use other procedures which are known to be equivalent in flat space–time.[11] For a discussion of the construction of Feynman propagators in curved space–time see Ref. 12.

The Feynman propagator may be written as the sum of real and imaginary parts

$$G_F(x,x') = -\tfrac{1}{2}\widetilde{D}(x,x') + \tfrac{1}{2}i\gamma(x,x'), \tag{5}$$

where γ is a real symmetric solution to Eq. (1). We will consider the various distributions G_F, $\widetilde{D}$, γ, and D to act on smooth test functions of compact support and we will always denote such test functions by f, g, and h.

When $D(x,x')$ acts on a test function f, it generates a solution of Eq. (1), since D is itself a solution of Eq. (1):

$$\phi_f(x) = \int D(x,x')f(x')d\tau(x'), \tag{6}$$

where $d\tau(x')$ is the covariant four volume element. Conversely, if we consider a solution $\phi(x)$ of Eq. (1) with compact spatial support, we can always find a test function $h(x)$ (not unique) such that

$$\int D(x,x')h(x')\,d\tau(x') = \phi(x). \tag{7}$$

To see this consider two spacelike hypersurfaces Σ_1 and Σ_2 with Σ_1 to the past of Σ_2. Define a new function $\hat{\phi}(x)$ by

 0022-2488/79/122506-05$01.00

setting $\hat{\phi}(x) = \phi(x)$ to the future of Σ_2 and $\hat{\phi}(x) \equiv 0$ to the past of Σ_1. Between Σ_2 and Σ_1, $\hat{\phi}$ is constructed to be smooth and match its values on Σ_1 and Σ_2. This can be done in many ways and is the source of the nonuniqueness. Finally, set

$$h(x) = (\Box - m^2)\hat{\phi}(x). \tag{8}$$

The $h(x)$ defined by Eq. (8) will reproduce $\phi(x)$ when substituted back in Eq. (6). This can be shown as follows: Let $h(x)$ be obtained from $\phi(x)$ by the above construction. Then we have for Eq. (6)

$$\phi_h(x) = \int D(x,x')(\Box' - m^2)\hat{\phi}(x')\, d\tau(x'), \tag{9}$$

where the prime on $\Box$ means that it acts on the variable x'. We now pick hypersurfaces Σ_f in the future and Σ_p in the past and integrate Eq. (9) by parts twice over a region bounded by Σ_f and Σ_p. The volume term will vanish since $D(x,x')$ is a solution of Eq. (1) and for the boundary term we get

$$\int_{\Sigma_f} \{\hat{\phi}(x')\nabla'^a D(x,x') - D(x,x')\nabla'^a\hat{\phi}(x')\} d\sigma_a(x'). \tag{10}$$

The integral over Σ_p is zero since we set $\hat{\phi} = 0$ in the past. In the future, however, $\hat{\phi}(x') = \phi(x')$, so it is obvious from Eq. (2) that the solution ϕ_h is the same as the ϕ we started with. We will denote solutions by ϕ_h, ϕ_g, etc. to denote that they correspond [via Eq. (6)] to test functions.

The action of the symplectic structure on solutions ϕ_g and ϕ_h can be expressed as a volume integral over the test functions g and h. We substitute for ϕ_g and ϕ_h in Eq. (3) using Eq. (6) to get

$$\Omega(\phi_g,\phi_h) = \int_\Sigma \left\{\left[\int D(x,x')g(x')\, d\tau(x')\right] \overset{\leftrightarrow}{\nabla}{}'^a \left[\int D(x,x'')h(x'')d\tau(x'')\right]\right\} d\sigma_a(x), \tag{11}$$

where the double arrow on ∇ is an abbreviation

$$a\overset{\leftrightarrow}{\nabla} b = a\nabla b - b\nabla a. \tag{12}$$

In Eq. (11) we interchange the surface and volume integrals

$$\Omega(\phi_g,\phi_h) = \int \left[\int_\Sigma \{D(x,x')\overset{\leftrightarrow}{\nabla}{}^a D(x,x'')\} d\sigma_a(x)\right] \times g(x')h(x'')d\tau(x')d\tau(x''). \tag{13}$$

However, since D is itself a solution of Eq. (1), it satisfies Eq. (2) so the surface integral in Eq. (13) reduces to $D(x',x'')$ and we get

$$\Omega(\phi_g,\phi_h) = \int D(x',x'')g(x')h(x'')d\tau(x')d\tau(x''). \tag{14}$$

This process of changing surface integrals into volume integrals will be frequently carried out by using Eqs. (2) and (6) and the fact that D and γ are solutions of Eq. (1).

A complex structure J acting on a real vector space V is a linear operator on V with the property that $J^2 = -1$. Such operators always exist on infinite dimensional spaces. Let V denote the space of real solutions to Eq. (1). Then if we have a decomposition of real solutions to Eq. (1) into positive and negative frequency parts $\phi^{(+)}$ and $\phi^{(-)}$, respectively, we can define a complex structure on V by

$$J\phi = i\phi^{(+)} - i\phi^{(-)}. \tag{15}$$

Note that $J\phi \in V$ even though $\phi^{(+)}$ and $\phi^{(-)}$ do not. $\phi^{(+)}$ and $\phi^{(-)}$ are complex solutions and belong to the complexified vector space $V_c = V' \oplus iV''$, where V' and V'' are copies of V. Conversely, if we have a complex structure J on V, we can define

$$P^+\phi = \phi^{(+)} = \frac{1}{2i}(i\phi + J\phi), \tag{16a}$$

$$P^-\phi = \phi^{(-)} = \frac{1}{2i}(i\phi - J\phi), \tag{16b}$$

where P^+ and P^- are maps from V into V_c. P^+ and P^- are the positive and negative frequency projection operators, respectively. It can be seen immediately that

$$P^{+2} = P^+, \quad P^{-2} = P^-, \quad \text{and}$$
$$P^+ + P^- = \text{identity}. \tag{17}$$

In stationary space–times there is a canonical complex structure.[3] In general, there is not. However, if the space–time has asymptotic static regimes in the past and future, we can define asymptotic complex structures in the past and future J_p and J_f, respectively. The fact that $J_p \neq J_f$ is what is responsible for particle creation.[3,4] Ashtekar and Magnon-Ashtekar have used the complex structure to define particle states and construct Fock spaces.[3,4] In their approach, they require that J and Ω be compatible in the sense that, for every real solutions ϕ of Eq. (1),

$$\Omega(\phi, J\phi) \geqslant 0. \tag{18}$$

This condition is imposed to ensure that the commutation relations between the field operators and the decomposition of a field operator into creation and annihilation operators are consistent.

We can define an operator

$$\mathscr{J} = (J_f - J_p)(J_f + J_p)^{-1}. \tag{19}$$

The existence of an S matrix relating the past and future Fock spaces depends on whether $\mathscr{J}$ is Hilbert–Schmidt.[13,14] The particle creation amplitudes can be given completely by $\mathscr{J}$ and vanish if $\mathscr{J}$ does. For details see Refs. 3–5 and 13.

III. RELATION BETWEEN PROPAGATOR AND COMPLEX STRUCTURE

We will first derive an expression for the Feynman propagator in terms of J_f and J_p. We will take the Feynman propagator to be given by

$$G_F(x,x') = \frac{i\langle \text{out} \mid T\phi(x)\phi(x') \mid \text{in}\rangle}{\langle \text{out} \mid \text{in}\rangle}. \tag{20}$$

This gives the usual Feynman propagator in flat space–time since $|\text{out}\rangle$ and $|\text{in}\rangle$ are the same there. In the formalism of Ref. 4 the J_f and J_p define $|\text{out}\rangle$ and $|\text{in}\rangle$, respectively. The quantum states are regarded as holomorphic functions on V, the space of classical solutions of Eq. (1). The usual Fock representation is recovered by taking the germ of the holomorphic function at the origin. The out vacuum state is the unit function while the image under S, of the in vacuum, is the function

$$S\,|\,\text{in}\rangle = K\exp(\tfrac{1}{2}\langle v, \mathscr{J}v\rangle_+), \tag{21}$$

where $v \in V$ and $\langle\ ,\ \rangle_+$ is the inner product in the future one

particle Hilbert space given by

$$\langle a,b\rangle_+ = \frac{1}{2h}\Omega(a,J_f b) + \frac{1}{2h} i\Omega(a,b). \tag{22}$$

The action of creation and annihilation operators $C(\phi)$ and $A(\phi)$, respectively, associated with a solution ϕ is

$$C(\phi)f(v) = \langle \phi,v\rangle_+ f(v), \tag{23a}$$

$$A(\phi)f(v) = \mathscr{L}_\phi f(v), \tag{23b}$$

where $v \in V$ and $f(v)$ is a holomorphic function on V.

We will calculate Eq. (20) with G_F "smeared out" with test functions h and g, where we have chosen h and g so that

$$\text{supp } h \cap \text{past of supp } g = \phi \tag{24}$$

to incorporate the time ordering. We then obtain for the smeared out Feynman propagator

$$\begin{aligned} G_F(h,g) &= \int G_F(x,x')h(x)g(x')d\tau(x)d\tau(x') \\ &= i\langle \text{out}|S|\text{in}\rangle^{-1}\langle \text{out}|[C(\phi_h)+A(\phi_h)] \\ &\quad \times [C(\phi_g)+A(\phi_g)]S|\text{in}\rangle. \end{aligned} \tag{25}$$

The inner products are taken in the future Fock space. Using the holomorphic function representation (21) for $S|\text{in}\rangle$ and Eq. (23) for the creation and annihilation operators,

$$\begin{aligned} G_F(h,g) = iK\langle 1,[C(\phi_h)+A(\phi_h)][C(\phi_g)+A(\phi_g)] \\ \times \exp\tfrac{1}{2}\langle v,\mathscr{J}v\rangle\rangle, \end{aligned} \tag{26}$$

where K is a constant. We note that only terms with annihilation opertors on the left survive [since $\langle \text{out} \mid C(\phi) = 0$] so we are left with

$$G_F(h,g) = iK\langle 1,[A(\phi_h)C(\phi_g)+A(\phi_h)A(\phi_g)]\exp\tfrac{1}{2}\langle v,v\rangle\rangle. \tag{27}$$

The first term is calculated easily using Eq. (23):

$$\begin{aligned} &A(\phi_h)C(\phi_g)\exp\tfrac{1}{2}\langle v,\mathscr{J}v\rangle \\ &\quad = \mathscr{L}_{\phi_h}(\langle \phi_g,v\rangle \exp\tfrac{1}{2}\langle v,\mathscr{J}v\rangle) \\ &\quad = \langle \phi_g,\phi_h\rangle \exp\tfrac{1}{2}\langle v,\mathscr{J}v\rangle + \langle \phi_g,v\rangle \mathscr{L}_{\phi_h}\exp\tfrac{1}{2}\langle v,\mathscr{J}v\rangle. \end{aligned} \tag{28}$$

Since we are taking the inner product with the constant term, only the first term contributes, giving a term equal to $\langle \phi_g,\phi_h\rangle$. Similarly, the second term can be shown to give rise to a term $\langle \phi_g,\mathscr{J}\phi_h\rangle$. There is a factor of $\langle \text{out}|S|\text{in}\rangle$ in both terms which cancels the factor $K = \langle \text{out}|S|\text{in}\rangle^{-1}$. Using the definition of the inner product (22), we get

$$\begin{aligned} G_F(h,g) = &-\tfrac{1}{2}\Omega[(1+\mathscr{J})\phi_g,\phi_h] \\ &+ \tfrac{1}{2}i\Omega[(1+\mathscr{J})\phi_g,J_f\phi_h]. \end{aligned} \tag{29}$$

Thus, we have expressed the Feynman propagator in terms of the complex structures and $\mathscr{J}$. Our subsequent analysis will consist of explicitly displaying the action of J_f, J_p, and $\mathscr{J}$ in terms of G_F. This will elucidate the relationship between the complex structure approach[1-4] and the propagator approach to field theory.

Consider the case where there is a unique complex structure J, i.e., $\mathscr{J} = 0$ and there is no particle creation. Then we obtain the following expression by setting $\mathscr{J} = 0$ in Eq. (29):

$$\begin{aligned} &\int G_F(x,x')h(x)g(x')d\tau(x)d\tau(x') \\ &\quad = -\tfrac{1}{2}\Omega(\phi_g,\phi_h) + \tfrac{1}{2}i\Omega(\phi_g,J\phi_h), \end{aligned} \tag{30}$$

where we have chosen g and h as in Eq. (24). Comparing Eqs. (30) and (5), we see that in this case

$$\int \gamma(x,x')h(x)g(x')d\tau(x)d\tau(x') = \Omega(\phi_g,J\phi_h). \tag{31}$$

We invert Eq. (31) to obtain J as follows: Pick a spacelike hypersurface Σ. Then define the action of J on ϕ by

$$J\phi(x) = \int_\Sigma \{\phi(x')\nabla'^a\gamma(x,x') - \gamma(x,x')\nabla'^a\phi(x')\}d\sigma_a(x'). \tag{32}$$

For J to be a complex structure we must have $J^2 = -1$. Imposing this condition and using Eq. (2), we obtain

$$\begin{aligned} &\int_\Sigma \{\gamma(x,y)\nabla_y^a\gamma(x',y) - \gamma(x',y)\nabla_y^a\gamma(x,y)\}d\sigma_a(y) \\ &\quad = D(x,x'). \end{aligned} \tag{33}$$

Thus, if γ is to define the action of J via Eq. (32), it must satisfy Eq. (33). We thus impose Eq. (33) as a condition on γ and are then assured that Eq. (32) defines a complex structure.

To obtain compatibility in the sense of Eq. (18) we demand that γ be positive definite in the sense that

$$\int \gamma(x,x')g(x)g(x')\,d\tau(x)\,d\tau(x') \geqslant 0 \tag{34}$$

for any real test function $g(x)$. It also follows that

$$\Omega(\phi_g,J\phi_h) = \Omega(\phi_h,J\phi_g) \tag{35}$$

from the symmetry of γ. By comparison of Eqs. (31) and (14) we see that if D and γ annihilate the same test functions (which they must since J is a linear operator), we must have the following condition:

$$\ker D \subset \ker\gamma. \tag{36}$$

The complex structure we have defined is clearly the same one as was implicit in Eq. (30), as can be seen by computing $\Omega(\phi_g,J\phi_h)$ using Eqs. (32) and (14) and the fact that γ satisfies Eq. (2).

We can express Eq. (32) as a four volume integral. If we use Eq. (6) in (32), we find that

$$\begin{aligned} J\phi_f(x) = &\int_\Sigma \left[\int D(x',y)f(y)d\tau(y)\right]\nabla'^a\gamma(x,x') \\ &- \gamma(x,x')\nabla'^a\left[\int D(x',y)f(y)d\tau(y)\right]d\sigma_a(x'), \end{aligned} \tag{37}$$

interchanging the orders of integration

$$\begin{aligned} J\phi_f(x) = \int\Big[\int_\Sigma &\{D(x',y)\nabla'^a\gamma(x,x') \\ &- \gamma(x,x')\nabla'^a D(x',y)\}d\sigma_a(x')\Big]f(y)d\tau(y). \end{aligned} \tag{38}$$

Since γ is itself a solution of Eq. (1), it satisfies Eq. (2) so we get

$$J\phi_f(x) = \int \gamma(x,y)f(y)d\tau(y). \tag{39}$$

The fact that this J is unique is guaranteed[13,15] if γ is Lie derived by the timelike Killing field. To see that the γ that we have produced is indeed the one which defines the vacuum in Eq. (20) we note that if we have some complex structure which defines a vacuum and we use Eq. (20) to define G_F, then 2 ImG_F is $\Omega(\phi_h, J\phi_g)$ by the argument following Eq. (20). If we use our definition of J and directly compute $\Omega(\phi_h, J\phi_g)$, a straightforward calculation reveals that we recover γ. Thus, the J we have defined is indeed the one implicit in γ.

We can summarize the situation as follows: A propagator G_F arises from a single complex structure J if and only if it satisfies in addition

(a) $\mathrm{Re}G_F(x,x') = \frac{1}{2}D(x,x')$, if $x \succ x'$,
$= -\frac{1}{2}D(x,x')$ if $x \prec x'$,

and zero if x and x' are spacelike related;

(b) $\gamma(x,x') \equiv 2\ \mathrm{Im}G_F(x,x')$ is symmetric in x and x' and satisfies Eq. (1);

(c) Ker $\gamma(x,x') = \mathrm{Ker}D(x,x')$;

(d) Eq. (33) is satisfied by $\gamma(x,x')$.

Condition (d) is stated by Lichnerowicz[2] and the existence of a γ satisfying these conditions is discussed by Moreno.[16]

We can obtain explicitly the positive and negative frequency parts of a solution $\phi_h(x)$ by using Eq. (16). One obtains the result (for x to the future of supp h but not in supp h).

$$\phi_h^{(+)}(x) = \int G_F(x,x')h(x')d\tau(x'), \tag{40a}$$

$$\phi_h^{(-)}(x) = \int G_F^*(x,x')h(x')d\tau(x'), \tag{40b}$$

where the * denotes complex conjugation. To obtain the solutions everywhere we use Eq. (2).

We now consider the case where there are two asymptotic complex structures[4] J_p and J_f. In this case there will be particle creation by the space–time geometry since $\mathscr{J} \neq 0$. Also, the expression for the propagator given by Eq. (29) is appropriate. This is the most general situation that can be described by an S matrix connecting particle states in the distant past and the distant future. We see immediately from Eq. (29) that if there is particle creation, the real part of G_F is not simply the symplectic structure. If we define

$$\tilde{D}(x,x') = -2\ \mathrm{Re}G_F(x,x'), \tag{41}$$

we can tell whether there is particle creation by comparing $\tilde{D}$ and D.

If there is particle creation, we can extract the action as follows: Define $\hat{D}$ as follows:

$$\hat{D}(x,x') = \tilde{D}(x,x') - D(x,x'). \tag{42}$$

where $\hat{D}$ acts on test functions g and h we obtain

$$\hat{D}(\phi_g,\phi_h) = \int \hat{D}(x,x')g(x)h(x')d\tau(x)d\tau(x')$$
$$= \Omega(\mathscr{J}\phi_g,\phi_h). \tag{43}$$

Let $\tilde{g}$ be a test function with the property

$$\int D(x,x')\tilde{g}(x')d\tau(x') = \mathscr{J}\phi_g. \tag{44}$$

Using Eq. (44) to rewrite the RHS of Eq. (43) in the form of Eq. (14),

$$\Omega(\mathscr{J}\phi_g,\phi_h) = \int D(x,x')\tilde{g}(x)h(x')d\tau(x)d\tau(x')$$
$$= \int \hat{D}(x,x')g(x)h(x')d\tau(x)d\tau(x'). \tag{45}$$

Since this is valid for arbitrary test functions h, we get

$$\mathscr{J}\phi_g(x) = \int \hat{D}(x,x')g(x')d\tau(x'). \tag{46}$$

For this to be well defined we must required

$$\mathrm{Ker}D \subset \mathrm{Ker}\hat{D}. \tag{47}$$

We now turn to the question of recovering J_f and J_p from G_F. The Feynman propagator defines the positive and negative frequencies in Eq. (40). In analogy with this we define

$$\phi_h^{(+)'}(x) = \int G_F(x,x')h(x')d\tau(x'), \tag{48a}$$

$$\phi_h^{(-)'}(x) = \int G_F^*(x,x')h(x')d\tau(x') \tag{48b}$$

to the future of but not including supp h. We obtain the solutions everywhere by choosing a spacelike hypersurface to the future of supp h and inducing the appropriate Cauchy data on it. We then use Eq. (2) to solve the Cauchy problem. Similarly, we obtain the past decomposition by defining

$$\phi_h^{(+)'}(x) = \int G_F(x,x')h(x')d\tau(x'), \tag{49a}$$

$$\phi_h^{(-)'}(x) = \int G_F^*(x,x')h(x')d\tau(x') \tag{49b}$$

in the past of but not including supp h. The Cauchy problem can again be used to obtain the solutions everywhere. We recover J_f and J_p by using Eq. (15). Thus,

$$J_f\phi_h(x) = \int \gamma(x,x')h(x')d\tau(x') \tag{50}$$

to the future of supp h, while to the past of supp h

$$J_p\phi_h(x) = \int \gamma(x,x')h(x')d\tau(x'). \tag{51}$$

These complex structures are not the same since the real part of G_F no longer governs the Cauchy evolution. Since we want $J_f\phi$ and $J_p\phi$ to be solutions of Eq. (1), we must demand that γ is a solution of Eq. (1). Then we can use Eq. (2) in Eqs. (50) and (51), and using Eq. (6) we can write the volume integrals as surface integrals to obtain

$$J_f\phi_h(x) = \int_\Sigma \{\phi_h(x')\nabla'^a\gamma(x,x')$$
$$- \gamma(x,x')\nabla'^a\phi_h(x')\}d\sigma_a(x') \tag{52a}$$

for Σ being a hypersurface to the future of supp h and x to the future of Σ. Similarly, we have

$$J_p\phi_h(x) = \int_\Sigma \{\phi_h(x')\nabla'^a\gamma(x,x')$$
$$- \gamma(x,x')\nabla'^a\phi_h(x')\}d\sigma_a(x') \tag{52b}$$

for Σ to the past of supp h and x to the past of Σ. We use these

forms to impose $J_p^2 = J_f^2 = -1$ and obtain convolution conditions exactly like Eq. (33). This is the case even though γ now involves as well as Ω and J. The formulas (48) and (49) can also be written in terms of surface integrals. We regard $G_F(x,x')$ as a function of x' for fixed x. This is a well behaved (nonsingular) solution of Eq. (1) for all $x' \neq x$. Thus, if we restrict x' to the future of x, we may use Eq. (2) in (48) or (49). We then perform the four volume integrals so that $\int D(x,x')h(x')d\tau(x')$ becomes $\phi_h(x)$ and obtain

$$\phi_h^{(+)1}(x) = \int_\Sigma \{G_F(x,x')\nabla'^a\phi_h(x') - \phi_h(x')\nabla'^a G_F(x,x')\}d\sigma_a(x'), \tag{53a}$$

$$\phi_h^{(+)1}(x) = \int_\Sigma \{G_F(x,x')\nabla'^a\phi_h(x') - \phi_h(x')\nabla'^a G_F(x,x')\}d\sigma_a(x'), \tag{53b}$$

for Σ being a hypersurface to the future of x. There is an analogous formula for the past decomposition. In this form the positive and negative frequency parts can be defined without reference to the test functions. Rumpf had earlier used these formulas to define positive and negative frequency parts.[8] His arguments used the analyticity properties of the propagator regarded as a function of m^2. As in the one complex structure case we require for consistency (a) $\mathrm{Ker}D \subset \mathrm{Ker}\hat{D}$, (b) γ is a symmetric solution of Eq. (1), and (c) γ obeys Eq. (33) on hypersurfaces in the distant past or distant future.

There are three principal approaches to quantum field theory in curved space–time: That of Lichnerowicz[2,16] based on Eq. (33), that of Segal,[1] Ashtekar,[3,4] and Kay[5] using complex structures, and that of DeWitt[7] and Rumpf[8] in which the propagator is the fundamental object. In this paper we have indicated the relations between these three approaches by explicitly displaying the complex structures in terms of the propagator and obtaining Eq. (33) as a necessary condition. We have also obtained conditions which a propagator must satisfy in order to qualify as being a legitimate propagator. Finally, since the propagator can often be explicitly calculated we can explicitly determine $\mathscr{J}$ and hence the S matrix, a calculation which is very difficult in the original formulation of Ashtekar and Magnon-Ashtekar[3,4] and Kay.[5]

ACKNOWLEDGMENTS

This research was supported in part by NSF grant #PH77-07111 and by a UWM Fellowship. I would like to thank John Friedman, A. Magnon-Ashtekar, A. Sen, R. Sorkin, and especially Abhay Ashtekar for helpful discussion. I would also like to thank L. Parker for advice and encouragement.

[1] I. Segal, J. Math. Phys. **1**, 468 (1960).
[2] A. Lichnerowicz, in *Relativity, Groups and Topology*, Les Houches, 1963. edited by B.S. DeWitt and C.M. DeWitt (Science, New York, 1964).
[3] A. Ashtekar and A. Magnon, Proc. R. Soc. (London) Ser. A **346**, 375 (1975).
[4] A. Ashtekar and A. Magnon-Ashtekar, "A Non-Perturbative Approach to Quantum Field Theory I," University of Chicago preprint (to be published).
[5] B.S. Kay, Commun. Math. Phys. **62**, 55 (1978).
[6] J. Schwinger, Proc. Natl. Acad. Sci. U.S.A. **37**, 452 (1951).
[7] B. DeWitt, Phys. Rep. C **19**, 295 (1975).
[8] H. Rumpf, Phys. Lett. B **61**, 272 (1976); and Ann. Phys. (Leipzig) **114**, 332 (1978).
[9] J.B. Hartle and S.W. Hawking, Phys. Rev. D **13**, 2188 (1976).
[10] D.M. Chitre and J.B. Hartle, Phys. Rev. D **16**, 251 (1977).
[11] B. DeWitt in *Relativity, Groups and Topology*, Les Houches, 1963, edited B.S. DeWitt and C.M. DeWitt (Science, New York, 1964).
[12] P. Candelas and D.J. Raine, Phys. Rev. D **15**, 1494 (1977).
[13] A. Magnon-Ashtekar, "Champs Quantiques en Espace–Temps Courbe," thesis submitted to Department de Mathematiques Pures, Universite de Clermont-Ferrand (unpublished), 1978.
[14] R.M. Wald, "On the Existence of the S-matrix," University of Chicago preprint, 1978.
[15] A. Ashtekar and A. Sen, "Quantum Fields in Curved Spacetime, Selection and Uniqueness of the Vacuum State," University of Chicago preprint, 1978.
[16] C. Moreno, J. Math. Phys. **18**, 2153 (1977).

Crossing properties of scattering operators

William H. Klink
Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242

(Received 25 May 1976; accepted for publication 5 June 1979)

For scattering operators that have the cluster decomposition property, it is shown that crossing relations for the scattering operator are consequences of the assumed crossing relations of the kernels of the connected parts of the scattering operator.

The goal of this paper is to derive crossing relations for scattering operators. The motivation for studying the crossing properties of scattering operators arises from trying to build an elementary particle theory in the spirit of the original Heisenberg theory,[1] in which the unitary scattering operator is constructed without the use of equations of motion. What we have in mind is a theory in which physical properties such as relativity, cluster properties, crossing and unitarity are expressed as operator relations.

The basic crossing operation to be used in this paper involves crossing one particle at a time,[2] either an initial or final particle (though for convenience most of the subsequent discussion will deal with crossing a final particle to an initial particle), in distinction to the more usual crossing operation, in which an initial and final particle are simultaneously crossed.[3] For multiparticle reactions crossing two particles simultaneously is equivalent to first crossing an initial (final) particle and then crossing a final (initial) particle. But for two particle to two particle reactions, the single particle crossing operation does not lead to a physical reaction.

To define crossing for scattering operators, we first consider the kernel of the connected scattering operator for an $A\rightarrow B$ reaction, where A is an initial and B a final cluster of particles. Now the kernel of the connected scattering operator can be written as the product of an energy momentum delta function times the connected scattering amplitude for the $A\rightarrow B$ reaction, which itself contains no further delta functions. As variables for this scattering amplitude, we will choose the energies E_i and directions $\hat{p}_i$ of the individual (incoming and outgoing) particles, for then all particles are treated on an equal footing. Such a choice of variables makes it easy to deal with identical particles where it is necessary to interchange particle labels. Because of the relativistic invariance of the scattering operator not all of the energies and angles are independent; but it is always possible to choose the energy of the particle being crossed as an independent variable. It is of course understood that all of the energies are positive. Along with energy and momentum variables, scattering amplitudes also depend on spin and internal symmetry variables. In the interests of simplicity such variables will be ignored in this paper. It is to be emphasized that it is not necessary to assume that the scattering operator is generated from some underlying Hamiltonian; instead general requirements such a relativistic invariance and unitarity are postulated.

The definition to be used for crossing is motivated by results from quantum field theory. If one particle (call it c) is crossed to become the antiparticle $\bar{c}$, there exists a path of analytic continuation for the scattering amplitude of the connected operator from the physical region of the direct channel where the energy E_c is positive to the physical region of the crossed channel, where it is negative. In this crossed physical region the analytic continuation of the scattering amplitude of the connected scattering operator for the direct reaction coincides with the scattering amplitude for the connected scattering operator of the crossed channel.

We begin the analysis of crossing properties of scattering operators by defining the projected or channel scattering operators. Consider an initial channel A of particles going to some final channel B of particles; not all of the particles need to interact with one another. Then the scattering operator for the $A\rightarrow B$ reaction can be written as

$$S_{B,A} = \Lambda_B S \Lambda_A, \tag{1}$$

where S is the scattering operator and Λ_A (Λ_B) is a projection operator from the full Fock space to the subspace $\mathscr{H}_A$ ($\mathscr{H}_B$).

The cluster properties for $S_{B,A}$ are physically motivated by requiring that if a cluster of particles is far away from another cluster, no interaction from the short-range hadronic forces should take place between the separated clusters. This physical requirement can be expressed mathematically in terms of strong operator limits of $S_{B,A}$.[4] If the particles in channel A are divided into two clusters A_1 and A_2, and the particles in channel B divided into clusters B_1 and B_2, then the cluster property can be expressed as

$$\text{strong} \lim_{|\mathbf{a}|\rightarrow\infty} U_{\mathbf{a}}^{\dagger} S_{B,A} U_{\mathbf{a}} = S_{B_1A_1} \otimes S_{B_2A_2}, \tag{2}$$

where $U_{\mathbf{a}} = \exp(-i\mathbf{P}_{A_1}\cdot\mathbf{a})$ and $\mathbf{P}_{A_1}$ is the momentum operator for cluster A_1. By letting A_1 and B_1 range over all different combinations of particles in A and B, respectively, a set of disconnected scattering operators for the $A\rightarrow B$ reaction is generated. Further sets of scattering operators can be generated by taking strong operator limits of the scattering operators of subclusters, until finally connected scattering operators are obtained.

Such connected scattering operators are characterized by the fact that

$$\text{strong} \lim_{|\mathbf{a}|\rightarrow\infty} U_{\mathbf{a}}^{\dagger} S_{B,A}^{(0)} U_{\mathbf{a}} = 0 \tag{3}$$

for all possible subclusters of A and B. That is, any scattering operator $S_{B,A}$ can be written as a sum of products of connected scattering operators, just because for a fixed cluster A of

 0022-2488/79/122511-03$01.00

initial particles, there are only a finite number of ways of dividing A into subclusters. This means there is a unique connected scattering operator $S^{(0)}_{B,A}$ associated with each scattering operator $S_{B,A}$.

Equation (3) states what is meant by a connected operator. The kernel of a connected operator $S^{(0)}_{B,A}$ can be written as $\delta^4(p_A - p_B)F^{B,A}$, where $F^{B,A}$ is the scattering amplitude for the $A \to B$ reaction and is analytic everywhere in the physical region except at physical region singularities. As stated in the Introduction, it is assumed that a path of analytic continuation for $F^{B,A}$ between the direct and crossed channels always exists.

To define a crossed connected operator let c denote the particle to be crossed and let A and B denote clusters as before. Then the kernel of the connected scattering operator for the $A + \bar{c} \to B$ reaction is related to the kernel of the connected scattering operator for the $A \to B + c$ reaction by the crossing relation

$$[\delta^4(p_A - p_B - p_c)F^{B+c,A}]^{cr} = \delta^4(p_A + p_{\bar{c}} - p_B)F^{B,A+\bar{c}}, \tag{4}$$

where $[\delta^4(p_A - p_B - p_c)]^{cr} = \delta^4(p_A - p_B + p_{\bar{c}})$ and $(F^{B+c,A})^{cr}$ is the analytic continuation of $F^{B+c,A}$ to the physical region of the $A + \bar{c} \to B$ reaction. To convert Eq. (4) to an operator equation, write

$$(S^{(0)}_{B,A+\bar{c}} f^{A+\bar{c}})\ (\mathbf{p}_j, j \in B)$$
$$= \int \prod_{i \in A+\bar{c}} \frac{d^3\mathbf{p}_i}{E_i} \delta^4(p_A + p_{\bar{c}} - p_B) F^{B,A+\bar{c}}(\mathbf{p}_j, \mathbf{p}_i) f^{A+\bar{c}}(\mathbf{p}_i),$$
$$(S^{(0)}_{B+c,A} f^A)\ (\mathbf{p}_j, j \in B + c)$$
$$= \int \prod_{i \in A} \frac{d^3\mathbf{p}_i}{E_i} \delta^4(p_A - p_c - p_B) F^{B+c,A}(\mathbf{p}_j, \mathbf{p}_i) f^A(\mathbf{p}_i), \tag{5}$$

where f^A is in $\mathscr{H}_A$ and $S^{(0)}_{B,A} f^A$ is in $\mathscr{H}_B$.

We define a new operator, the crossed connected scattering operator, by

$$S^{(0)}_{B+c,A}(cr) f^{A+\bar{c}}$$
$$\equiv \int \prod_{i \in A+\bar{c}} \frac{d^3\mathbf{p}_i}{E_i} [\delta^4(p_A - p_c - p_B) F^{B+c,A}]^{cr} f^{A+\bar{c}}. \tag{6}$$

From this definition and the crossing relation, Eq. (4), a crossing relation between connected operators for $A \to B + c$ and $A + \bar{c} \to B$ reaction immediately follows:

$$S^{(0)}_{B+c,A}(cr) = S^{(0)}_{B,A+\bar{c}}. \tag{7}$$

If there are particles in cluster B that are identical with particle c, it is clear thar $F^{B+c,A}$ must be properly symmetrized with respect to the identical particles. Further, though the action of the connected operators in Eq. (5) is given with respect to momentum variables of the individual particles other choices of variables could also be made. In particular Ref. 2 gives a set of variables in which the total momentum of A, along with the invariant masses s_A of cluster A and s_B of cluster B are used; these variables have the important property that the analytic continuation of $F^{B+c,A}$ involves only an analytic continuation in s_B and s_A, with all other variables held fixed. Such types of variables will be of special interest in the succeeding paper where crossing relations are combined with unitarity.

To proceed from a connected crossed scattering operator to a full crossed scattering operator, it is necessary to be somewhat more explicit about the form of the cluster decomposition of $S_{B+c,A}$. Once particle c is chosen it follows from Eqs. (2) and (3) that it is possible to write all the disconnected terms in the cluster decomposition of $S_{B+c,A}$ as tensor products of connected operators involving particle c times remaining operators not involving c:

$$S_{B+c,A} = S^{(0)}_{c,c} \otimes S_{B,A-c} + \sum_{j,A_1(j)} S^{(0)}_{c+j,A_1(j)} \otimes S_{B-j,A_2}$$
$$+ \sum_{j,k,A_1(j,k)} S^{(0)}_{c+j+k,A_1(j,k)} \otimes S_{B-j-k,A_2}$$
$$+ \cdots + S^{(0)}_{B+c,A}, \tag{8}$$

with $A_1 \cup A_2 = A$.

The definition of $S_{B+c,A}(cr)$ will be given in terms of the crossed operators in the cluster decomposition of $S_{B+c,A}$. Thus it is necessary to extend the definition of a crossed operator from connected operators to the disconnected operators occurring in the cluster decomposition of $S_{B+c,A}$. Such an extension is straightforward because all of the operators in Eq. (8) involving particle c are connected. A typical term in Eq. (8) can be written as $S^{(0)}_{B_1+c,A_1} \otimes S_{B-B_1,A-A_1}$, where A_1 and B_1 are subclusters of A and B, respectively. Now

$$(S^{(0)}_{B_1+c,A_1} \otimes S_{B-B_1,A-A_1}) f^A$$
$$= \int \prod_{i \in A} \frac{d^3\mathbf{p}_i}{E_i} [\delta^4(p_{A_1} - p_c - p_{B_1}) F^{B_1+c,A_1}$$
$$\times \delta^4(p_{A-A_1} - p_{B-B_1}) K^{B-B_1,A-A_1}] f^A \tag{9}$$

so that a crossed disconnected operator can be defined as

$$(S^{(0)}_{B_1+c,A_1} \otimes S_{B-B_1,A-A_1})(cr)$$
$$\equiv S^{(0)}_{B_1+c,A_1}(cr) \otimes S_{B-B_1,A-A_1}. \tag{10}$$

From this definition of a crossed disconnected operator, a crossing relation immediately follows:

$$(S^{(0)}_{B_1+c,A_1} \otimes S_{B-B_1,A-A_1})(cr)$$
$$= S^{(0)}_{B_1,A_1+\bar{c}} \otimes S_{B-B_1,A-A_1}, \tag{11}$$

where $S^{(0)}_{B_1,A_1+\bar{c}} \otimes S_{B-B_1,A-A_1}$ is a disconnected operator occurring in the cluster decomposition of $S_{B,A+\bar{c}}$. $K^{B-B_1,A-A_1}$ in Eq. (9) is the kernel of the operator $S_{B-B_1,A-A_1}$ which itself can be written as a tensor product of connected operators—that is, as a product of momentum delta functions and connected F amplitudes. Some care must be exercised in using the crossing relation, Eq. (11), for the right-hand side may be zero depending on the particles in cluster B_1.

For example the simplest operator, the "straight thru" or unit operator in which none of the particles interact, can be written

$$S^{\text{straight thru}}_{B+c,A} f^A = f^A \tag{12}$$

so the kernel consists only of delta functions. In particular the delta function $\delta^4(p'_c - p_c)$, where p'_c is the initial and p_c the final four momentum of particle c, indicates that the four momentum of particle c is unchanged. Crossing the "straight thru" operator then gives

$$S^{\text{straight thru}}_{B+c,A}(cr)f^{A+\bar{c}} = \int d^4p'_c \cdots \delta^4(p_c + p'_{\bar{c}}) \cdots f^{A+\bar{c}}. \quad (13)$$

This example shows that the energy momentum delta functions may force some of the disconnected terms to be zero under crossing, for there is no physical region where $\delta^4(p_c + p'_{\bar{c}})$ is nonzero for positive energy positive mass particles.

In fact it is easy to isolate those disconnected operators that give zero contribution under crossing by looking at the delta functions $\delta^4(p_{A_i} - p_c - p_{B_i})$ in Eq. (7). If B_1 is the cluster containing either no particles (as was the case for the "straight thru" operator) or the cluster containing only one particle, then after being crossed the delta function $\delta^4(p_{A_i} + p_{\bar{c}} - p_{B_i})$ will always give zero contribution to the cluster decomposition of $S_{B,A+\bar{c}}$.

Because of such delta functions there will not be a crossing relation of the type $S_{B+c,A}(cr) = S_{B,A+\bar{c}}$, as was the case for connected operators, Eq. (7) [and most disconnected operators, Eq. (11)]; rather terms must be added in order to relate $S_{B+c,A}(cr)$ to $S_{B,A+\bar{c}}$. But the analysis of the previous paragraphs shows that it is just the one line and $N \to 2$, $N \geqslant 2$ connected pieces of a general disconnected term that need to be considered. By suitably adding in these terms it is possible to relate $S_{B,A+\bar{c}}$ to $S_{B+c,A}(cr)$ plus disconnected operators in which the one line and $N \to 2$, $N \geqslant 2$ connected operators involve particle c, the particle being crossed. If $S_{B+c,A}(cr)$ is defined through the cluster decomposition of $S_{B+c,A}$, Eq. (8), namely

$$\begin{aligned} S_{B+c,A}(cr) \equiv S^{(0)}_{c,c}(cr) \otimes S_{B,A-c} + \sum_{j,A_1(j)} S^{(0)}_{c+j,A_1(j)}(cr) \\ \otimes S_{B-j,A_2} + \sum_{j,k,A_1(j,k)} S^{(0)}_{c+j+k,A_1(j,k)}(cr) \\ \otimes S_{B-j-k,A_2} + \cdots + S^{(0)}_{B+c,A}(cr), \end{aligned} \quad (14)$$

then the crossing relation can be written as

$$\begin{aligned} S_{B,A+\bar{c}} = S_{\bar{c},\bar{c}} \otimes S_{B-\bar{c},A} + \sum_{j,B_1(j)} S^{(0)}_{B_1(j),\bar{c}+j} \\ \otimes S_{B-B_1(j),A-j} + S_{B+c,A}(cr). \end{aligned} \quad (15)$$

Equation (15) is the desired crossing relation; it relates the S operator for the $A + \bar{c} \to B$ reaction to the crossed S operator for the $A \to B + c$ reaction plus disconnected operators involving certain connected subreactions for $\bar{c}$. These subreactions are either of the from $\bar{c} \to \bar{c}$ ($\bar{c}$ does not interact with other particles) or $\bar{c} + j \to B_1(j)$, in which $\bar{c}$ interacts with any particle j from cluster A to produce a final cluster $B_1(j)$. An expression similar to that of Eq. (15) can be written for $S_{B+c,A}$ in terms of $S_{B,A+\bar{c}}(cr)$ plus one line and $N \to 2$ type connected subreactions involving particle c. It is then straightforward to check that if the expression for $S_{B,A+\bar{c}}$ given in Eq. (15) is substituted into the analogous expression for $S_{B+c,A}$, an identity in $S_{B+c,A}$ results. In obtaining this identity care must be taken in applying "(cr)" twice. The operation "cr" is defined only after the particle to be crossed has been chosen. To apply the crossing operation twice, that is, first to particle c and then to particle $\bar{c}$ means using the cluster decompositions [Eq. (8)] appropriate to particles c and $\bar{c}$. Only if these cluster decompositions are used will "cr" applied first to c and then to $\bar{c}$ result in an identity in $S_{B+c,A}$.

In concluding it is necessary to ask how identical particles might affect the crossing relation [Eq. (15)]. In the cluster decomposition, Eq. (8), it was (implicitly) assumed that each disconnected term could be written as though the particles were distinguishable. However for particles in B that are identical to c, it is necessary to check that the terms are all properly symmetrized. This is most easily done when the cluster decomposition of S is given in terms of connected scattering operators only. Then there are two ways in which a particle c' identical to c can be related to c. Either c' occurs in the same connected amplitude as c or it does not. If c' and c are in the same connected amplitude, then the connected amplitude must be appropriately symmetrized with respect to c and c'. If c' is not in the same connected amplitude as c then, because of the cluster decomposition of S into connected parts, there must be a term with c and c' interchanged. That is, for the cluster decomposition of S into connected parts, proper symmetrization is assured if each connected amplitude containing identical particles is properly symmetrized. Then the terms in such a cluster decomposition can be grouped according to Eq. (8) and crossing carried out as though there were no particles identical with c.

We have shown how to obtain crossing relations not only for connected, but also for disconnected and full scattering operators. In the following paper,[5] these crossing relations will be used in conjunction with unitarity equations to represent scattering operators so that crossing and unitarity are automatically satisfied.

ACKNOWLEDGMENTS

The author wishes to thank Professor Fritz Coester for many helpful discussions.

[1] W. Heisenberg, Z. Physik **120**, 513 (1942), 678 (1943).
[2] W.H. Klink, Nucl. Phys. B **86**, 175 (1975).
[3] See, for example, G.F. Chew, *The Analytic S Matrix* (Benjamin, New York, 1966) or R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The Analytic S Matrix* (Cambridge University, Cambridge, England, 1966).
[4] F. Coester, Helv. Phys. Acta **38**, 7 (1965).
[5] W.H. Klink, J. Math. Phys., see following paper.

A representation of multiparticle scattering operators satisfying unitarity and crossing properties

W. H. Klink
Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242

(Received 25 May 1977; accepted for publication 5 June 1979)

A representation for multiparticle scattering operators satisfying unitarity and crossing properties is presented. The representation is given in terms of a set of functions that satisfy orthogonality, completeness and analyticity conditions. It is shown that integrals over these functions yield inclusive cross sections.

The framework for an S matrix theory was first proposed by Heisenberg in 1943.[1] His idea was to constrain the possible forms the S matrix could have by making use of relativity and unitarity. This was done by writing $S = e^{i\eta}$, where η is Hermitian[2] and then choosing various forms for η in order to make predictions for cross sections of production reactions. In the spirit of Heisenberg's original proposal, we wish to construct a representation for multiparticle scattering operators that not only automatically satisfies relativity and unitarity, but also crossing relations.

As discussed in the previous paper,[3] the theory under consideration is basically an operator theory in that physical properties such as a cluster decomposition or unitarity are stated as operator relations on the scattering operator. The previous paper discussed how relations between crossed reactions could be stated as operator relations between scattering operators. Also cluster properties can be stated in terms of strong operator limits of scattering operators.[4] In this paper unitarity[5] will be analyzed and it will be shown how a spectral representation for the scattering operator can be obtained, which, when combined with crossing relations, allows one to express all multiparticle operators as products of a set of functions that must satisfy certain orthogonality, completeness, and analyticity conditions. That is, this set of functions will automatically satisfy unitarity and crossing relations. To indicate the close connection of this set of functions with quantities of physical interest, the paper concludes by showing that there is a simple relationship between integrals of these basic functions and inclusive cross sections.

In order to focus on the basic features of the representaton, we will only deal with reactions of the form two nucleons plus any number of pions reacting to produce two nucleons plus any (other) number of pions. Spin and internal symmetries such as isospin will not be considered; generalizations to arbitrary numbers of nucleons with spin and internal symmetries will be discussed in subsequent papers. Between the three and four-body thresholds there are then four reactions that are related by unitarity, namely, $NN \to NN$, $NN \to NN\pi$, $NN\pi \to NN$, and $NN\pi \to NN\pi$. The manner in which the scattering operators for these reactions are related can be obtained by projecting the unitarity equations $SS^\dagger = S^\dagger S = I$ into the various (initial and final) two and three-particle subspaces of the Fock space. For the four reactions mentioned above, this gives $S_{B,A} = \Lambda_B S \Lambda_A$ where Λ_A and Λ_B are projection operators into either a two-particle (NN) or three-particle ($NN\pi$) subspace. Because of the relativistic invariance of S it is most convenient to write these two and three-particle subspaces as direct integrals of two and three-particle partial wave Hilbert spaces. More generally the projection of Fock space to an n-particle Hilbert space can be written as the direct integral

$$\int d^{\,3}\mathbf{P} \int ds \sum_{J,\sigma} \oplus \mathscr{H}_{\mathbf{P}sJ\sigma;n},$$

where $\mathbf{P}$ is the total momentum, $\sqrt{s}$ the "mass," J the total angular momentum, and σ the spin projection of the n-particle system. $\mathscr{H}_{\mathbf{P}sJ\sigma;n}$ is the n-particle partial wave Hilbert space. Since we will be dealing almost exclusively with partial wave Hilbert spaces in this paper, the subscripts $\mathbf{P}$, s, J, and σ will be omitted and the n particle partial wave Hilbert space written as $\mathscr{H}_n$.

For example, a two-particle wave function may be written in terms of the individual momenta of the two particles; however, a change of variables to the total momentum $\mathbf{P}$, the two-particle mass $\sqrt{s} = \sqrt{(p_1 + p_2)^2}$, and J, σ obtained by Legendre transforming the direction of the relative two-particle momentum allows one to specify the action of $S_{2,2} = \Lambda_2 S \Lambda_2$ as

$$S_{2,2} f = \eta(sJ)\, e^{2i\delta(sJ)} f, \quad f \in \mathscr{H}_2 \tag{1}$$

where η is (up to a phase space factor) the inelasticity parameter and δ the phase shift. Actually it is possible to drop the element f in Eq. (1) since the two-particle partial wave space $\mathscr{H}_2$ is one dimensional.

Similarly $S_{3,2}$ is an operator from $\mathscr{H}_2$ to $\mathscr{H}_3$, the three-particle partial wave space. If f is in $\mathscr{H}_2$ (i.e., is a number), then

$$S_{3,2} f = \mathscr{A}^{2\to 3} f, \quad S_{3,2} = \mathscr{A}^{2\to 3}, \tag{2}$$

where $\mathscr{A}^{2\to 3}$ is the $NN \to NN\pi$ partial wave amplitude. Equation (2) states that the operator $S_{3,2} = \Lambda_3 S \Lambda_2$ can be viewed as a vector in $\mathscr{H}_3$.

The unitarity equations for two-particle initial and final states can thus be written as

$$\begin{aligned} &\Lambda_2 S^\dagger \Lambda_2 \Lambda_2 S \Lambda_2 + \Lambda_2 S^\dagger \Lambda_3 \Lambda_3 S \Lambda_2 = \Lambda_2, \\ &\eta^2 + \| \mathscr{A}^{2\to 3} \|^2 = 1, \\ &\Lambda_2 S \Lambda_2 \Lambda_2 S^\dagger \Lambda_2 + \Lambda_2 S \Lambda_3 \Lambda_3 S^\dagger \Lambda_2 = \Lambda_2, \\ &\eta^2 + \| \mathscr{A}^{3\to 2} \|^2 = 1. \end{aligned} \tag{3}$$

 0022-2488/79/122514-06$01.00

Here $\mathscr{A}^{3\to 2}$ is related to the operator $S_{2,3} = \Lambda_2 S \Lambda_3$ by

$$S_{2,3} f = (\mathscr{A}^{3\to 2}, f) \in \mathscr{H}_2, \quad f \in \mathscr{H}_3 \tag{4}$$

where $\mathscr{A}^{3\to 2}$ is also an element in $\mathscr{H}_3$. Two-particle unitarity, Eq. (3), thus states that the lengths of the $2\to 3$ and $3\to 2$ partial wave amplitudes are equal.

There are other unitarity equations relating the four operators. The most improtant ones, resulting from three-particle initial and final states, can be written

$$\begin{aligned} &\Lambda_3 S \Lambda_2 \Lambda_2 S^\dagger \Lambda_3 + \Lambda_3 S \Lambda_3 \Lambda_3 S^\dagger \Lambda_3 = \Lambda_3, \\ &S_{3,2} S_{3,2}^\dagger + S_{3,3} S_{3,3}^\dagger = \Lambda_3, \\ &I - S_{3,3} S_{3,3}^\dagger = \mathscr{A}^{2\to 3} \otimes \mathscr{A}^{2\to 3\dagger} \ \forall f \in \mathscr{H}_3. \end{aligned} \tag{5a}$$

Here I is the unit operator in $\mathscr{H}_3$ and $S_{3,2}^\dagger$ is defined as $(\Lambda_3 S \Lambda_2)^\dagger$ from which it follows that $S_{3,2}^\dagger f = (\mathscr{A}^{2\to 3}, f) \in \mathscr{H}_2$, for f in $\mathscr{H}_3$. Thus, $S_{3,2} S_{3,2}^\dagger f = S_{3,2}(\mathscr{A}^{2\to 3}, f) = \mathscr{A}^{2\to 3}(\mathscr{A}^{2\to 3}, f) \equiv (\mathscr{A}^{2\to 3} \otimes \mathscr{A}^{2\to 3\dagger}) f$. A similar equation follows from $S^\dagger S = I$:

$$\begin{aligned} &\Lambda_3 S^\dagger \Lambda_2 \Lambda_2 S \Lambda_3 + \Lambda_3 S^\dagger \Lambda_3 \Lambda_3 S \Lambda_3 = \Lambda_3, \\ &S_{2,3}^\dagger S_{2,3} + S_{3,3}^\dagger S_{3,3} = \Lambda_3, \\ &I - S_{3,3}^\dagger S_{3,3} = \mathscr{A}^{3\to 2} \otimes \mathscr{A}^{3\to 2\dagger} \ \forall f \in \mathscr{H}_3, \end{aligned} \tag{5b}$$

where the adjoint of $S_{2,3}$, $S_{2,3}^\dagger = (\Lambda_2 S \Lambda_3)^\dagger$, satisfies $S_{2,3}^\dagger = \mathscr{A}^{3\to 2}$. The last lines of Eqs. (5a) and (5b) indicate that the $3\to 3$ operator misses being a unitarity operator by the existence of the production partial wave amplitudes.

We would like to use Eq. (5) to spectrally represent $S_{3,3}$. But as can be seen from Eq. (5), $S_{3,3}$ is not a normal operator. An associated normal operator $\widetilde{S}$ can be defined by making use of the fact that $||\mathscr{A}^{3\to 2}|| = ||\mathscr{A}^{2\to 3}||$ and defining a unitary operator W from $\mathscr{H}_3$ to $\mathscr{H}_3$ by $\mathscr{A}^{3\to 2} = W \mathscr{A}^{2\to 3}$ (W is obviously highly nonunique). Then if $\widetilde{S}$ is defined by $\widetilde{S} = S_{3,3} W$, Eq. (5), written in terms of $\widetilde{S}$ shows that $\widetilde{S}$ is bounded and normal; that is, $[\widetilde{S}, \widetilde{S}^\dagger] f = 0$ for all f in $\mathscr{H}_3$. A spectral representation can then be written as

$$\begin{aligned} \widetilde{S} &= \int d\mu(\gamma)\, \lambda_\gamma e_\gamma \otimes e_\gamma^\dagger, \\ S_{3,3} &= \int d\mu(\gamma) \lambda_\gamma e_\gamma \otimes d_\gamma^\dagger, \end{aligned} \tag{6}$$

where γ stands for the set of labels needed to specify the eigenvectors, $d\mu(\gamma)$ is the unknown spectral measure and the $\{e_\gamma\}$ form a complete orthonormal set in $\mathscr{H}_3$. d_γ is defined as $W e_\gamma$, and since W is unitary the set $\{d_\gamma\}$ also is complete and orthonormal in $\mathscr{H}_3$. Though the $\{d_\gamma\}$ are defined by the operator W, since nothing at this point is known about W except that $\mathscr{A}^{3\to 2} = W \mathscr{A}^{2\to 3}$, we will regard the $\{d_\gamma\}$ as a complete set of vectors that are independent of the $\{e_\gamma\}$ and make use of further physical requirements to relate the $\{d_\gamma\}$ and $\{e_\gamma\}$.

By applying Eq. (5a) to a basis element e_γ, it is readily seen that the generalized eigenvalues λ_γ all have modulus 1 except for one eigenvalue λ_1 for which $|\lambda_1| = \eta$; here $\gamma = 1$ has been chosen to designate that eigenvector for which $\mathscr{A}^{2\to 3} = \sqrt{(1-\eta^2)} e_1$ (similarly $\mathscr{A}^{3\to 2} = \sqrt{(1-\eta^2)} \times d_1$). Aside from the fact that e_1 is a true eigenvector of $\widetilde{S}$, nothing is known about the spectrum of $\widetilde{S}$. The sense in which the right-hand side of Eq. (6) represents $S_{3,3}$ is in terms of a strong operator limit.[6]

To generalize these unitarity results to arbitrary numbers of pions,[7] it suffices to replace $\mathscr{H}_3$ by the direct sum

$$\mathscr{H}' = \mathscr{H}_3 \oplus \mathscr{H}_4 \oplus \cdots \oplus \mathscr{H}_N, \tag{7}$$

where N is the maximum number of particles allowed by the fixed mass $\sqrt{s}$. Define projection operators Λ_2 and Λ', where Λ_2 projects into the one-dimensional two-particle partial wave space as before, and Λ' projects the full Fock space into the direct sum space $\mathscr{H}'$. Then the unitarity equations will have the same form as the three-particle unitarity equations, Eq. (5). In particular, each operator $S_{B,A} = \Lambda_B S \Lambda_A$, where A and B are channels containing one or more pions, can be written

$$S_{B,A} = \int d\mu(\gamma) \lambda_\gamma e_\gamma^B \otimes d_\gamma^{A\dagger}, \tag{8}$$

where $e_\gamma^B = \Lambda_B e_\gamma$ and $d_\gamma^B = \Lambda_B W e_\gamma$; for a fixed channel B some of the e_γ^B and d_γ^B may of course be zero.

The projected vectors e_γ^B and d_γ^B are not in general orthogonal; only their direct sum is orthogonal:

$$\begin{aligned} &\sum_B (e_\gamma^B, e_{\gamma'}^B) = \delta_{\gamma\gamma'}, \\ &\sum_B (d_\gamma^B, d_{\gamma'}^B) = \delta_{\gamma\gamma'}, \end{aligned} \tag{9}$$

where the sum over B means a sum over all the partial wave Hilbert spaces in the direct sum Hilbert space $\mathscr{H}'$. Thus the $\{e_\gamma^B\}$ and $\{d_\gamma^B\}$ each span $\mathscr{H}_B$, but do not in general form orthonormal bases in $\mathscr{H}_B$. As was the case for three-particle unitarity, $|\lambda_\gamma| = 1$ for $\gamma \neq 1$. We will fix the "eigenvectors" e_γ and d_γ, $\gamma \neq 1$, so that $\lambda_\gamma = 1$; that is, e_γ and d_γ will absorb the phases coming from λ_γ, $\gamma \neq 1$. As was the case for three-particle unitarity the production partial wave amplitudes are given by $\mathscr{A}^{2\to B} = \sqrt{(1-\eta^2)} e_1^B$ and $\mathscr{A}^{B\to 2} = \sqrt{(1-\eta^2)} d_1^B$, with $|\lambda_1| = \eta$ the inelasticity parameter. The argument of λ_1 is related to the two-particle to two-particle phase shift.

Thus far the spectral representation for the channel operators $S_{B,A}$ has not referred to any particular set of variables; to prepare for an analysis of crossing, it is convenient to choose a set of variables that refers specifically to the particle (call it c) being crossed. As in Ref. 3 crossing refers to the transformation of one incoming (outgoing) particle to an outgoing (incoming) antiparticle. The more usual use of the term "crossing" to designate the transformation of two particles, one incoming and one outgoing, to their respective antiparticles is equivalent to two single particle crossing operations for multiparticle reactions. If variables for elements of $\mathscr{H}_B$ are used as variables for elements of $\mathscr{H}_{B+c}$, it is possible to expand the eigenfunctions e_γ^{B+c} and d_γ^{B+c} in terms of e_γ^B and d_γ^B, respectively:

$$\begin{aligned} e_\gamma^{B+c} &= \int d\mu(\gamma') e_{\gamma'}^B f_{\gamma'\gamma}^{B+c}, \\ d_\gamma^{B+c} &= \int d\mu(\gamma')\, d_{\gamma'}^B\, g_{\gamma'\gamma}^{B+c}, \end{aligned} \tag{10}$$

where $f_{\gamma'\gamma}^{B+c}$ and $g_{\gamma'\gamma}^{B+c}$ may be thought of as expansion coeffi-

cients that contain variables from elements of $\mathscr{H}_{B+c}$ not occurring in elements of $\mathscr{H}_B$.

To be more specific about these variables, we choose a definite ordering of the particles in the B cluster, say 1,2,...,j with 1,2 always referring to the two nucleons. Then the individual momenta of the B cluster particles can be replaced by the unit vector directions of all the B cluster particles, along with subcluster masses defined by $s_i = (p_1 + \cdots + p_i)^2$, $i = 2...j$. The phase space measure $d^3\mathbf{p}_1/E_1 \cdots d^3\mathbf{p}_j/E_j$ can be written in these new variables as[8]

$$dw = d\hat{p}_j \frac{|\mathbf{p}_j|}{\sqrt{s_j}} ds_{j-1} \cdots d\hat{p}_3 \frac{|\mathbf{p}_3|}{\sqrt{s_2}} ds_2 \, |\mathbf{p}_2|, \tag{11}$$

where $|\mathbf{p}_i| = [\lambda^{1/2}(s_i, s_{i-1}, m_i)]/(2\sqrt{s_i})$, $\lambda(x,y,z) = x^2 + y^2 + z^2 - 2(xy + xz + yz)$ is the magnitude of the momentum of the ith particle in the frame where $\mathbf{p}_1 + \cdots + \mathbf{p}_i = \mathbf{0}$ and the limits of integration of the cluster masses $\sqrt{s_i}$ is from $m_1 + \cdots + m_{i-1}$ to $\sqrt{s_{i+1}} - m_i$ (m_l is the mass of the l th particle). $d\hat{p}_i = d(\cos\theta_i)\, d\varphi_i$ where the angles θ_i and φ_i are chosen as

$\cos\theta_i = \hat{p}_i \cdot \hat{p}_{i+1}$ in the frame where $\mathbf{p}_1 + \cdots + \mathbf{p}_i = \mathbf{0}$,

$\varphi_i =$ azimuthal angle between the planes defined by

$$\mathbf{p}_i - \mathbf{p}_{i+1} \text{ and } \mathbf{p}_{i+1} - \mathbf{p}_{i+2}, \text{ also in the frame where}$$

$$\mathbf{p}_1 + \cdots + \mathbf{p}_i = \mathbf{0}, \tag{12}$$

because then they have simple crossing properities.[9]

Since we are interested in partial wave amplitudes, the angles θ_i and φ_i are Legendre transformed to J_i, the angular momentum of the 1,2,...,i cluster and σ_i, the spin projection along an axis defined by the $i+1$ particle.[9] In these variables the phase space volume becomes

$$\int dw = \int_{(m_1 + \cdots + m_{j-1})^2}^{(\sqrt{s_j} - m_j)^2} \frac{|\mathbf{p}_j|}{\sqrt{s_j}} ds_{j-1} \cdots \int_{(m_1+m_2)^2}^{(\sqrt{s_3} - m_3)^2} \frac{|\mathbf{p}_3|}{\sqrt{s_3}} \times ds_2 \, |\mathbf{p}_2| \sum_{J_j \sigma_j} \cdots \sum_{J_2 \sigma_2}. \tag{13}$$

The cluster variables just defined can be used in the partial wave Hilbert spaces $\mathscr{H}_B$; however, in these spaces J_B and σ_B are fixed parameters, so that when computing an inner product of elements in $\mathscr{H}_B$, there is no sum over J_B and σ_B [i.e., over J_j and σ_j in Eq. (13)].

It is now possible to use the cluster variables for elements in $\mathscr{H}_B$ and $\mathscr{H}_{B+c}$ and thus give a precise meaning to the expansion, Eq. (10). In particular, e_γ^{B+c} can be written in terms of cluster variables as $e_\gamma^{B+c}(s_B\, J_B \sigma_B \cdots)$, where the dots refer to further cluster variables and the superscript $B+c$ refers not only to the $B+c$ cluster, but also to the (fixed) labels s_{B+c} and J_{B+c}. Now e_γ^B does not depend on σ_B and depends parametrically on s_B and J_B. Thus, e_γ^{B+c} is an element of $\mathscr{H}_B$ for $s_B\, J_B \sigma_B$ held fixed. So the expansion coefficients $f_{\gamma\gamma'}^{B+c}$ will have as variables $s_B\, J_B \sigma_B$ and depend parametrically on s_{B+c} and J_{B+c}. Since the Hilbert space $\mathscr{H}_B$ is nested in $\mathscr{H}_{B+c}$, it is possible to view the expansion coefficients as elements of a Hilbert space with a norm given by

$$||f_{\gamma\gamma'}^{B+c}||^2 = \sum_{J_B \sigma_B} \int_{m_B^2}^{(\sqrt{s_{B+c}} - m_{B+c})^2} \frac{|\mathbf{p}_c|}{\sqrt{s_{B+c}}} ds_B \times |f_{\gamma\gamma'}^{B+c}(s_{B+c}\, J_{B+c}; s_B\, J_B \sigma_B)|^2 < \infty, \tag{14}$$

where m_B is the sum of the masses of all the particles in cluster B.

In Ref. 3 a crossed scattering operator was defined by first defining a crossed *connected* scattering operator through the analytic continuation of its associated scattering amplitude. The operator $S_{B,A+\bar{c}}(cr)$ was then defined as the sum of all the crossed scattering operators that occur in the cluster decomposition of $S_{B,A+\bar{c}}$. Using the functions $g_{\gamma\gamma'}$, the spectral representation of $S_{B,A+\bar{c}}$ can be written as

$$S_{B,A+\bar{c}} = \int d\mu(\gamma) \lambda_\gamma e_\gamma^B \otimes d_\gamma^{A+\bar{c}\dagger} = \int d\mu(\gamma)\, d\mu(\gamma') \lambda_\gamma e_\gamma^B \otimes (d_{\gamma'}^A\, g_{\gamma'\gamma}^{A+\bar{c}})^\dagger. \tag{15}$$

We now make the fundamental assumption that if particle c is crossed, the crossed channel operator $S_{B,A+\bar{c}}(cr)$ can be represented as

$$S_{B,A+\bar{c}}(cr) = (-1)^{J_A + J_B} \int d\mu(\gamma)\, d\mu(\gamma') \lambda_\gamma e_\gamma^B \times g_{\gamma'\gamma}^{A+\bar{c}(cr)*} \otimes d_{\gamma'}^{A\dagger}, \tag{16}$$

where (cr) denotes the analytic continuation of $g_{\gamma'\gamma}^{A+\bar{c}}$ in the variables s_A and s_B. That is, we assume that $g_{\gamma'\gamma}$ (and $f_{\gamma'\gamma}$) have analytic properties that allow them to be analytically continued in the variables s_A and s_B from the physical region of the $A + \bar{c} \to B$ channel where $\sqrt{s_B} = \sqrt{s_{A+\bar{c}}} \geqslant \sqrt{s_A} + m_c$ to the physical region of the $A \to B + c$ channel where $\sqrt{s_A} \geqslant \sqrt{s_B} + m_c$. The term $(-1)^{J_A + J_B}$ arises because there is not only an analytic continuation in s_A and s_B but also an interchanging of roles of J_A, the angular momentum of the A cluster, which is also the total angular momentum of the $A \to B + c$ reaction, with J_B, the angular momentum of the B cluster, which is also the angular momentum of the $A + \bar{c} \to B$ reaction. Under crossing the momentum vector of particle c changes direction which—in partial wave variables—gives the factor $(-1)^{J_A + J_B}$. Reference 9 presents these arguments in greater detail, and also shows that under crossing the spin projection variable σ of cluster B becomes the spin projection variable of cluster $B + c$. Thus, $g_{\gamma'\gamma}^{A+\bar{c}(cr)}(s_B\, J_B; s_A\, J_A \sigma)$ designates an analytic continuation in the variables s_B and s_A to the physical region where $\sqrt{s_A} \geqslant \sqrt{s_B} + m_c$, while J_A becomes the total angular momentum of the $A \to B + c$ reaction and J_B is the angular momentum of the B cluster.

On the other hand, it is possible to represent $S_{B+c,A}$ as

$$S_{B+c,A} = \int d\mu(\gamma') \lambda_{\gamma'} e_{\gamma'}^{B+c} \otimes d_{\gamma'}^{A\dagger} = \int d\mu(\gamma')\, d\mu(\gamma) \lambda_\gamma e_\gamma^B f_{\gamma\gamma'}^{B+c} \otimes d_{\gamma'}^{A\dagger}, \tag{17}$$

where in contrast to $g_{\gamma\gamma'}$, the variable dependence of the $f_{\gamma\gamma'}$ is $f_{\gamma\gamma'}^{B+c}(s_A\, J_A; s_B\, J_B \sigma)$, with $\sqrt{s_A} \geqslant \sqrt{s_B} + m_c$. By comparing the representation of $S_{B,A+\bar{c}}(cr)$ [Eq. (16)] with the re-

presentation for $S_{B+c,A}$ given in Eq. (17), it is seen that $S_{B,A+\bar{c}}(cr)$—like $S_{B+c,A}$—carries elements from $\mathscr{H}_A$ to $\mathscr{H}_{B+c}$. Thus, crossing a scattering operator means carrying out an analytic continuation on the $f_{\gamma\gamma'}^{B+c}$ or $g_{\gamma\gamma'}^{A+\bar{c}}$ functions. In order for such an operation to be compatible with the spectral representation of the scattering operators—where the final channel variables are not functionally related with the initial channel variables, we assume that the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ do not depend on the number of particles in cluster B or cluster A. Then the superscript $B+c$ on the $f_{\gamma\gamma'}$ and $A+\bar{c}$ on the $g_{\gamma\gamma'}$ can be removed and the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ become "universal" expansion coefficients, valid for any number of particles. In the following paragraphs it will be shown that the functions $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ are not independent of one another, but can be related through the use of crossing relations that connect $S_{B+c,A}$ with $S_{B,A+\bar{c}}(cr)$. Also, crossing relations can be used to express the e_γ^3 and d_γ^3 in terms of the $g_{\gamma\gamma'}$ and $f_{\gamma\gamma'}$ functions. Since e_γ^B and d_γ^B can be expanded with respect to the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$, respectively, it is seen that all channel scattering operators can be expressed in terms of the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$.

Before showing how crossing relations can be used to express the e_γ^B and d_γ^B in terms of the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ functions, we want to demonstrate that the universal character of the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ is consistent with the cluster properties of the scattering operators. For simplicity choose cluster A to be a three-particle cluster. Reference 3 shows that the crossing relation then becomes

$$S_{B+c,A} = S_{B,A+\bar{c}}(cr) + S_{c,c} \otimes S_{B,A-c} \tag{18}$$

which in terms of the spectral representation gives

$$\begin{aligned}\int d\mu(\gamma)\, d\mu(\gamma')\lambda_{\gamma'}(A)e_\gamma^B f_{\gamma\gamma'} \otimes d_{\gamma'}^{A\dagger} &= (-1)^{J_A+J_B}\int d\mu(\gamma)\, d\mu(\gamma')\lambda_{\gamma'}(B)e_\gamma^B \otimes [d_{\gamma'}^A g_{\gamma'\gamma}{}^{cr}]^\dagger \\ &+ \mathscr{J}^{-1}\delta^3[B-(A-c)]\sqrt{1-\eta^2}\,e_1^B,\end{aligned} \tag{19}$$

where $\lambda_\gamma(A)$ means $\lambda_\gamma(s_A\ J_A)$ and $\delta^3(B-A)$ designates the three delta functions $\delta(s_B - s_A)\delta_{J_BJ_A}\delta_{\sigma_B\sigma_A}$; $\mathscr{J}$ is the Jacobian of Eq. (14). Taking the inner product of Eq. (19) from the left with respect to $e_{\gamma''}^B$ and summing over the particle number B (while holding s_B and J_B fixed) gives—using the orthogonality relations, Eq. (9)—

$$\begin{aligned}\int d\mu(\gamma')f_{\gamma\gamma'}\lambda_{\gamma'}(A) \otimes d_{\gamma'}^{A\dagger} &= (-1)^{J_A+J_B}\int d\mu(\gamma')\lambda_{\gamma'}(B)g_{\gamma'\gamma}{}^{(cr)*} \\ &\otimes d_{\gamma'}^{A\dagger} + \mathscr{J}^{-1}\delta^3[B-(A-c)]\sqrt{1-\eta^2(B)}\ \delta_{\gamma 1}.\end{aligned} \tag{20}$$

Now $f_{\gamma\gamma'}$ is an element of a Hilbert space with a norm given by Eq. (14). On the right-hand side of Eq. (20) there are also elements of this Hilbert space, but they do not depend on the B cluster. Therefore, we conclude that $f_{\gamma\gamma'}$ can be chosen to be independent of the number of particles in cluster B.

A further consequence of the crossing relation, Eq. (18) is obtained when the B cluster is chosen to consist of two particles. Then e_γ^3 and d_γ^3 can be shown to also depend on the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$. If the A cluster is chosen to consist of more than three particles, Eq. (18) can be spectrally represented as

$$\begin{aligned}\int d\mu(\gamma)\lambda_\gamma(A)e_\gamma^3 \otimes d_\gamma^{A\dagger} &= (-1)^{J_A+J_B}\sqrt{1-\eta^2(B)}\int d\mu(\gamma)(d_\gamma^A g_{\gamma 1}^{(cr)})^\dagger \\ &+ \mathscr{J}^{-1}\delta^3[B-(A-c)]\sqrt{1-\eta^2(B)}\ d_1^{A-c\dagger}.\end{aligned} \tag{21}$$

To express e_γ^3 in terms of $g_{\gamma\gamma'}$, we write d_γ^A as $\int d\mu(\gamma')\times d_{\gamma'}^{A-c}g_{\gamma'\gamma}$ and make use of the fact that d_γ^{A-c} spans the direct sum Hilbert space summed over A to get

$$\begin{aligned}\int d\mu(\gamma)\left[\lambda_\gamma(A)e_\gamma^3 - (-1)^{J_A+J_B}\sqrt{1-\eta^2(B)}\ g_{\gamma 1}^{*(cr)}\right]g_{\gamma'\gamma}^* &= \sqrt{1-\eta^2(B)}\ \mathscr{J}^{-1}\delta^3[B-(A-c)]\,\delta_{\gamma'1}.\end{aligned} \tag{22}$$

To actually express e_γ^3 in terms of $g_{\gamma\gamma'}$ it is necessary to investigate how the orthogonality properties of $e_\gamma^B(d_\gamma^B)$ determine orthogonality properties of $f_{\gamma\gamma'}(g_{\gamma\gamma'})$. When $\sqrt{s_B}$ is between the three and four-body threshold, only e_γ^B, where B is a three-particle cluster, can contribute in the orthogonality relation, Eq. (9). Then

$$(e_\gamma^3, e_{\gamma'}^3) = \delta_{\gamma\gamma'}, \tag{23}$$

for $\sqrt{s_B}$ between the three and four-body threshold. Between the four and five-body threshold the orthogonality relations, Eq. (9) read

$$(e_\gamma^4, e_{\gamma'}^4) + (e_\gamma^3, e_{\gamma'}^3) = \delta_{\gamma\gamma'}. \tag{24}$$

If e_γ^4 is expanded as $\int d\mu(\bar{\gamma})e_{\bar{\gamma}}^3 f_{\bar{\gamma}\gamma}$, the invariant mass variable appearing in the $e_{\bar{\gamma}}^3$ must be between the three and four-body threshold, so that the $(e_{\bar{\gamma}}^3, e_{\bar{\gamma}'}^3)$ appearing in the expansion of $(e_\gamma^4, e_{\gamma'}^4)$ of Eq. (24) will give a $\delta_{\bar{\gamma}\bar{\gamma}'}$, according to Eq. (23). Therefore,

$$\int d\mu(\bar{\gamma})(f_{\bar{\gamma}\gamma}, f_{\bar{\gamma}\gamma'}) + (e_\gamma^3, e_{\gamma'}^3) = \delta_{\gamma\gamma'}, \tag{25}$$

where the inner product $(f_{\bar{\gamma}\gamma}, f_{\bar{\gamma}\gamma'})$ is defined in Eq. (14). Though Eq. (25) holds only for $\sqrt{s}$ between the four and five-body thresholds, we will assume that it holds for all energies above the four-body threshold. For then it is not difficult (but rather tedious) to show that the orthogonality relations of Eq. (9) hold for any energy. Basically all that is required is to show that if Eq. (9) is valid for B clusters consisting of 3,4,..., j numbers of particles, from Eq. (25) it follows that Eq. (9) holds for B clusters consisting of 3,4,..., $j, j+1$ particles. Thus, the orthogonality relations of Eq. (9) will automatically be satisfied if we assume that Eq. (25) is valid for all energies above the four-particle threshold; there is of course a similar expression for the $g_{\gamma\gamma'}$ and d_γ^3.

Using the orthogonality relation, Eq. (25), it is possible to simplify the crossing relation, Eq. (22). To that end the spectral representation of Eq. (18) when B is a two-particle cluster and A is three-particle cluster can be written as

$$\int d\mu(\gamma)\left[\lambda_\gamma(A)e_\gamma^3 - (-1)^{J_A+J_B}\sqrt{1-\eta^2(B)}\ g_{\gamma 1}^{*(cr)}\right] \otimes d_\gamma^3$$

$$= \mathscr{I}^{-1}\delta^3\,[B-(A-c)]\;\mathscr{A}^{2\to 2}(B), \tag{26}$$

where $\mathscr{A}^{2\to 2}$ is the 2→2 partical partial wave amplitude. Equations (22) and (24) can then be written as

$$\int d\mu(\gamma)\,\left[\lambda_\gamma(A)e^3_\gamma - (-1)^{J_A+J_B}\sqrt{1-\eta^2(B)}\;g^{*(cr)}_{\gamma 1}\right] \times \sum_{\gamma'}(g_{\gamma\gamma'}, g_{\gamma\bar{\gamma}}) + (d^{\frac{3}{\gamma}}, d^{\frac{3}{\bar{\gamma}}}) = \sqrt{1-\eta^2(B)}\;g_{1\bar{\gamma}} + \mathscr{A}^{2\to 2}(B)d^{\frac{3}{\bar{\gamma}}},$$

and using the orthogonality relation, Eq. (24) gives

$$\lambda_{\bar{\gamma}}(A)e^3_{\bar{\gamma}} = \sqrt{1-\eta^2(B)}\;\left[(-1)^{J_A+J_B}g^{*(cr)}_{\gamma 1} + g_{1\bar{\gamma}}\right] + \mathscr{A}^{2\to 2}\,d^{\frac{3}{\bar{\gamma}}}. \tag{27}$$

An analogous expression can be written for $d^{\frac{3}{\gamma}}$ in terms of $f^{cr}_{\gamma 1}, f_{1\gamma}$, and e^3_γ. Taken together such equations show that e^3_γ and $d^{\frac{3}{\gamma}}$ can be written in terms of $f_{\gamma\gamma'}$, $g_{\gamma\gamma'}$ and their respective crossed functions.

It is not the goal in this paper to write out the consequences of all the crossing relations explicitly. The reason is that when both the A and B particle clusters contain more than three particles the crossing relation, Eq. (18) is not complete. Rather, as shown in Ref. 3, two-particle disconnected terms must be added which make the spectral representation considerably more complicated. What is missing in the crossing relation, Eq. (18)—as shown in Ref. 3—are all the two-line disconnected terms of the form $S_{c+l,A_1}\otimes S_{B-l,A_2}$ with $A_1\cup A_2 = A$. If such disconnected operators are spectrally represented, it is possible to show that the $f_{\gamma\gamma'}$ appearing in Eq. (19) can be written in terms of the $g_{\gamma\gamma'}$, where $g_{\gamma\gamma'}$ arise not only from $S_{B,A+\bar{c}}(cr)$ and $S_{c,c}\otimes S_{B,A-c}$ as before, but also from the two-line disconnected terms given above. What makes these two-line disconnected terms complicated to handle is they connot be naturally expressed in terms of the cluster variables of Eq. (13). Rather the natural cluster variables that are used to spectrally represent each channel scattering operator in the two-line disconnected terms must be transformed with the help of Poincaré group Racah coefficients into the correct cluster variables. Before these two-line disconneted terms can be properly handled, it is necessary to work out these Racah coefficients for n particles, as generalizations of the three-particle Racah coefficients.[10]

Thus, unitarity and crossing leads to a representation of the channel scattering operators in terms of a basic set of related functions $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ that satisfy orthogonality relations of the type given in Eq. (25) and are independent of cluster size.

As an application of these ideas, we conclude by showing that the e^B_γ (and hence the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$) are closely tied to inclusive cross sections. Consider at first only energies up to the four-body threshold. Using crossing relations the 2→4 partial wave amplitude can be written

$$\mathscr{A}^{2\to 4} = (-1)^{J_A+J_B}\int d\mu(\gamma)\lambda_\gamma(B)\,e^{(B)}_\gamma\,d^{(A+\bar{c})*cr}_\gamma, \tag{28}$$

where A refers to the initial two-particle (NN) cluster, and B refers to the outgoing three-particle ($NN\pi$) cluster. To get an inclusive cross section we first Legendre transform both sides of Eq. (28) by $P_{J_A\sigma}(\theta)$ to convert the total angular momentum J_A to an angle θ, the angle between the incoming particles and outgoing detected pion (in the overall CM). Then Eq. (28) becomes

$$\tilde{\mathscr{A}}^{2\to 4} = (-1)^{J_B}\int d\mu(\gamma)\eta_\gamma(B)e^{(B)}_\gamma\tilde{d}^{(A+\bar{c})cr*}, \tag{29}$$

where

$$\tilde{d}^{(A+\bar{c})cr}_\gamma = \sum_{J_A} P_{J_A\sigma}\,d^{(A+\bar{c})cr}_\gamma(-1)^{J_A}.$$

To get an inclusive cross section, we write

$$\left(\int\sum\right)_{B'}|\tilde{\mathscr{A}}^{2\to 4}|^2 = \int d\mu(\gamma)\,d\mu(\gamma')\lambda_\gamma\tilde{d}^{(A+\bar{c})cr*}_\gamma(e^{(B)}_\gamma, e^{(B)}_{\gamma'})\lambda^*_{\gamma'}\tilde{d}^{(A+\bar{c})cr}_{\gamma'} = \int d\mu(\gamma)\,|\lambda_\gamma|^2\,|\tilde{d}^{(A+\bar{c})cr}_\gamma|^2 = \eta^2\,|\tilde{d}^{(A+\bar{c})cr}_1|^2 + \int_{\gamma\neq 1} d\mu(\gamma)\,|\tilde{d}^{(A+\bar{c})cr}_\gamma|^2 = \int d\mu(\gamma)\,|\tilde{d}^{(A+\bar{c})cr}|^2 - (1-\eta^2)\,|\tilde{d}^{(A+\bar{c})cr}_1|^2, \tag{30}$$

where B' designates the labels for the outgoing two-particle NN cluster. Rewriting Eq. (30) then gives

$$\int d\mu(\gamma)\,|\tilde{d}^{(A+\bar{c})cr}_\gamma|^2 = |\sqrt{1-\eta^2}\,\tilde{d}^{(A+\bar{c})cr}_1|^2 + \left(\int\sum\right)_{B'}|\tilde{\mathscr{A}}^{2\to 4}|^2 = |(-1)^{J_B}\tilde{\mathscr{A}}^{2\to 3}|^2 + \left(\int\sum\right)_{B'}|\tilde{\mathscr{A}}^{2\to 4}|^2, \tag{31}$$

where use has been made of the fact that $(\mathscr{A}^{3\to 2})^{cr} = (-1)^{J_A+J_B}\mathscr{A}^{2\to 3}$. $\tilde{\mathscr{A}}^{2\to 3}$ is the Legendre transform of the $\mathscr{A}^{2\to 3}$ partial wave amplitude. Now the right-hand side of Eq. (31) can be converted to 2→3 and 2→4 amplitudes by transforming with respect to the spherical harmonics $Y_{J_B\sigma}(\Omega)$. In fact, using the completeness of the spherical harmonics and summing over J_B and σ gives

$$\sum_{J_B\sigma}\int d\mu(\gamma)\,|\tilde{d}^{(A+\bar{c})cr}_\gamma|^2 = \int d\Omega\left[|A^{2\to 3}|^2 + \left(\int\sum\right)_{B'}|A^{2\to 4}|^2\right] = \sigma_{\text{inclusive}}(s_A, \cos\theta, s_B). \tag{32}$$

That is, by taking suitable sums over the d_γ functions one obtains an inclusive cross section. But the d_γ functions are related to the basic $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ functions. Thus, sums of the $f_{\gamma\gamma'}$ functions lead directly to an inclusive cross section.

To generalize this result to arbitrary energies it is simply necessary to make use of the spectral representation for the $A+\bar{c}\to B$ operator, where again A is a two-particle cluster, but now B contains more than one pion. Equation (18) gives

$$S_{B+c,A} = S_{B,A+\bar{c}}(cr) = (-1)^{J_A+J_B}\int d\mu(\gamma)\lambda_\gamma e^{(B)}_\gamma\otimes d^{(A+\bar{c})cr\dagger}_\gamma, \tag{33}$$

$$\mathscr{A}^{A\to B+C} = (-1)^{J_1+J_{II}} \int d\mu(\gamma)\, \lambda_\gamma e_\gamma^{(B)}\, d_\gamma^{(A+\bar{c})cr^*},$$

where the last line of Eq. (33) follows from the fact that $S_{B+C,A}$ acts on elements in $\mathscr{H}_A$, which is a one-dimensional space. Again, if both sides of Eq. (33) are Legendre transformed with respect to $P_{J_1\sigma}(\theta)$, the equation can be written

$$\tilde{\mathscr{A}}^{A\to B+C} = (-1)^{J_{II}} \int d\mu(\gamma)\lambda_\gamma e_\gamma^{(B)}\, \tilde{d}_\gamma^{(A+\bar{c})cr^*},$$

where $\tilde{d}_\gamma$ is defined as before [see Eq. (29) ff]. Now, however, the $e_\gamma^{(B)}$ are not orthonormal for a fixed number of particles in B. Rather it is necessary to sum over all possible clusters, which means summing over all possible numbers of pions that can be produced for a given initial energy s_A:

$$\sum_{\substack{B \text{ cluster} \\ \text{variables}}} \left(\int \Sigma\right) |\tilde{\mathscr{A}}^{A\to B+C}|^2$$

$$= \int d\mu(\gamma)\, d\mu(\gamma')\lambda_\gamma\lambda_{\gamma'}\,(e_\gamma^{(B)}, e_{\gamma'}^{(B)})\, \tilde{d}_\gamma^{(A+\bar{c})cr^*}\, \tilde{d}_{\gamma'}^{(A+\bar{c})cr}$$

$$= \int d\mu(\gamma)\left|\tilde{d}_\gamma^{(A+\bar{c})cr}\right|^2 - \sqrt{1-\eta^2}\,\tilde{d}_1^{(A+\bar{c})cr}\Big|^2,$$

$$\int d\mu(\gamma)\left|\tilde{d}_\gamma^{(A+\bar{c})cr}\right|^2$$

$$= |\tilde{\mathscr{A}}^{2\to 3}|^2 + \sum_{\substack{B \text{ cluster} \\ \text{variables}}} \left(\int \Sigma\right) |\tilde{\mathscr{A}}^{A\to B+C}|^2. \tag{34}$$

Finally if the $\tilde{\mathscr{A}}$ partial wave amplitudes are converted to amplitudes, the result becomes identical with Eq. (32), now valid at all energies.

$$\sum_{J_{II}\sigma} \int d\mu(\gamma)\left|\tilde{d}_\gamma^{(A+\bar{c})cr}\right|^2 = \sigma_{\text{inclusive}}(s_A, \cos\theta, s_B). \tag{35}$$

Equation (35) states that the "eigenvectors" of $S_{B,A+\bar{c}}$, when converted into amplitudes and suitably summed and integrated, give the inclusive cross section at all energies with respect to the detected pi meson, that is, the particle that as crossed. A similar result can be shown to hold for two particle inclusive cross sections by suitably crossing two particles using the representation for the $S_{B,A+\bar{c}}$ operator where now A is a three-particle cluster, and in general the N particle inclusive cross section, resulting from crossing $N-2$ particles in the A particle cluster.

We have shown how a basic set of functions can be used to represent multiparticle scattering operators in such a way as to guarantee that the scattering operators will satisfy unitarity and crossing relations. This has of course only been demonstrated on a restricted class of production reactions (no $N\bar{N}$ pair production) and for particles with no spin or internal symmetries. Nevertheless, it is clear how such sets of functions—the $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$—could be broadened to include more realistic particle reactions.

In stressing the importance of this basic set of functions, it is important to be clear as to the assumptions that have been made that lead to the conclusions of this paper. The $f_{\gamma\gamma'}$ and $g_{\gamma\gamma'}$ are assumed to not only have orthogonality properties of the type given in Eq. (25), but also have analyticity properties, so that they can be analytically continued to crossed channels. These analyticity properties are postulated to give meaning to the most important assumption of this paper, namely that the spectral representation of a channel scattering operator can be analytically continued to a crossed channel to represent the crossed channel scattering operator. It is not at all clear how such an assumption might be checked. Rather the usefulness of such an assumption is given by the degree to which it is possible to actually construct models of multiparticle scattering operators that satisfy general physical requirements, including not only those discussed in this paper, but also other requirements such as time reversal invariance and causality.

Along with these assumptions, there are also further requirements that must be investigated in order to develop a sensible operator S matrix theory. Thus far it has been assumed that the nucleons are inert in the sense that they cannot be crossed or form $N\bar{N}$ pairs. Allowing the nucleons to be crossed means dealing with internal symmetries and conserved changes, so the set of reactions that are allowed and connected by crossing and unitarity become larger and more complicated. Further it is necessary to let the particles have intrinsic spin and impose the correct statistics on the multiparticle amplitudes.

In spite of the fact that these issues have not yet been dealt with, it is our contention that the $f_{\gamma\gamma'}$ functions offer a suitable starting point for investigating multiparticle processes. For by approximating the spectral measure by a finite sum, it is possible to construct model multiparticle amplitudes that automatically satisfy relativity, crossing, unitarity and cluster properties. There are of course other properties that would not automatically be satisfied by these model amplitudes, for example, causality. So, on the one other hand, by building in more properties such as causality or time reversal invariance, it should be possible to restrict the class of $f_{\gamma\gamma'}$ functions and perhaps even be able to fix the spectral measure. On the other hand, it should be possible to approximate the $f_{\gamma\gamma'}$ functions with functions having finite spectral measure, and in this way construct actual phenomenological models for multiparticle reactions. Both of these possibilities are being explored.

[1]W. Heisenberg, Z. Physik **120**, 513 (1942); **121** 678 (1943).
[2]I. Bialynicki-Birula, B. Mielnik, and J. Plebański, Ann. Phys. (N.Y.) **51**, 187 (1969).
[3]W.H. Klink, "Crossing Properties of Scattering Operators," J. Math. Phys. (1978).
[4]F. Coester, Helv. Phys. Acta **38**, 7 (1965).
[5]W.H. Klink, Nucl. Phys. **B77**, 56 (1974).
[6]See, for example, P.R. Halmos, *Introduction to Hilbert Space and the Theory of Spectral Multiplicity* (Chelsea, New York, 1951), Chap. II or N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963), Vol. II, Chap. X.
[7]In this paper the number of nucleons is fixed and no $N\bar{N}$ pair production is allowed. If pair production and crossing of nucleons were allowed, sets of unitarity equations with different generalized charges would result, making the representation more complicated. Such matters will be taken up in subsequent papers.
[8]E. Byckling and K. Kajantie, *Particle Kinematics* (Wiley, New York, 1973).
[9]W.H. Klink, Nucl. Phys. **B 86**, 175 (1975).
[10]W.H. Klink, J. Math. Phys. **16**, 1247 (1975).

Off-shell scattering by Coulomb-like potentials

H. van Haeringen

Natuurkundig Laboratorium der Vrije Universiteit, Amsterdam, The Netherlands
and Institute for Theoretical Physics, P. O. Box 800, University of Groningen, The Netherlands

(Received 5 September 1978)

We derive closed expressions for and interrelationships between off-shell and on-shell scattering quantities for Coulomb plus short-range potentials. In particular we introduce off-shell Jost states and show how the transition matrices are obtained from these states. We discuss some formulas connecting the coordinate and momentum representatives of certain quantities. For the pure Coulomb case we derive analytic expressions for the Jost state and the off-shell Jost state for $l = 0$ in the momentum representation.

1. INTRODUCTION

In this paper we study off-shell scattering by a potential which is the sum of the Coulomb potential and a local central potential of short range. We derive many interesting expressions, notably for the Jost functions, the off-shell Jost functions, and the on-shell and off-shell "Jost states." These quantities are closely connected with the transition matrix which plays such an important role in scattering theory.

First, in Sec. 2, we confine ourselves to a general local short-range central potential. Here we derive many interrelationships between the above quantities. Only a few of these are well known, e.g., the defining expression for the (off-shell) Jost functions in terms of the (off-shell) Jost solutions in the coordinate representation. We give the momentum representation equivalents of these expressions which have a somewhat simpler form.

Some of the equations given in Sec. 2 are also valid for Coulomblike potentials. However, some have to be modified for such potentials with a long range. To this end we consider in Sec. 3 the pure Coulomb case. By working out a number of explicit expressions we pave the way for the treatment of the general case of Coulomb plus shortrange potentials, which will be given in Sec. 4. We also prove the validity of two conjectures made in Ref. 1.

Furthermore, in Sec. 3 we derive some interesting analytic expressions, notably for the $l = 0$ Coulomb Jost state and the off-shell Jost state in the momentum representation. In these expressions we encounter a certain hypergeometric function which appears in many other Coulomb quantities. Only its argument is different for the various different cases.

We will use mainly the notation of Refs. 1 and 2. In particular the energy is k^2 with $\mathrm{Im}k \downarrow 0$ and the energy dependence of G, G_0 and T will be suppressed. However, instead of the Jost solution $f_l(k,r)$ and the off-shell Jost solution $f_l(k,q,r)$ of the *radial* differential equations we will use the Jost solution $\langle r|kl\uparrow\rangle$ and the off-shell Jost solution $\langle r|kql\uparrow\rangle$ of the partial-wave projected equations. Here q is an off-shell momentum variable for which we assume $\mathrm{Im}q \geqslant 0$. We shall also consider the Hankel transforms of the above Jost solutions. These are denoted by $\langle p|kl\uparrow\rangle$ and $\langle p|kql\uparrow\rangle$, respectively. We call $|kl\uparrow\rangle$ the Jost state and $|kql\uparrow\rangle$ the off-shell Jost state.

2. THE SHORT-RANGE POTENTIAL CASE

In this section we confine ourselves to a local central potential $V(r)$, having a short range. Let us first recall Fuda's definition of the off-shell Jost solution[3]: $f_l(k,q,r)$ is that solution of the inhomogeneous differential equation

$$\left(k^2 + \frac{d^2}{dr^2} - \frac{l(l+1)}{r^2} - V(r)\right) f_l(k,q,r) = (k^2 - q^2) i^l q r h_l^{(+)}(qr), \tag{2.1}$$

which satisfies the asymptotic condition

$$\lim_{r\to\infty} f_l(k,q,r) e^{-iqr} = 1.$$

We introduce the "state" $|kql\uparrow\rangle$ by

$$\langle r|kql\uparrow\rangle \equiv (2/\pi)^{1/2}(qr)^{-1} f_l(k,q,r). \tag{2.2}$$

This may be compared with the "state" $|kl\uparrow\rangle$ that we introduced before,

$$\langle r|kl\uparrow\rangle \equiv (2/\pi)^{1/2}(kr)^{-1} f_l(k,r). \tag{2.3}$$

Let $H_l = H_{0l} + V_l$ be the partial-wave Hamiltonian, then we obtain from Eq. (2.1),

$$(k^2 - H_l)|kql\uparrow\rangle = (k^2 - q^2)|ql\uparrow\rangle_0, \tag{2.4a}$$

that is,

$$G_l^{-1}|kql\uparrow\rangle = G_{0l}^{-1}|ql\uparrow\rangle_0. \tag{2.4b}$$

Here $|ql\uparrow\rangle_0$ is the Jost state corresponding to $V \equiv 0$. In the coordinate representation one has

$$\langle r|ql\uparrow\rangle_0 = (2/\pi)^{1/2} i^l h_l^{(+)}(qr).$$

Furthermore, we denote the scattering state for $V \equiv 0$ and energy k^2 by $|kl\rangle$, or by $|k\rangle$ when no confusion arises, e.g., $T_l|kl\rangle = T_l|k\rangle$. It should be noted that Eqs. (2.4) are valid only in the coordinate representation. We shall call $\langle r|kql\uparrow\rangle$ the off-shell Jost solution of the "inhomogeneous Schrödinger equation" corresponding to Eqs. (2.4).

We would like to have a closed expression for $|kql\uparrow\rangle$. It is easily seen from Eq. (2.4b) that $G_l G_{0l}^{-1}|ql\uparrow\rangle_0$ is a particular solution of an *in*homogeneous differential equation. If one adds to this quantity an arbitrary solution of the corresponding homogeneous differential equation it remains a solution of (2.4). Now we have, again in the coordinate representation only,

$$(k^2 - H_l)|kl\uparrow\rangle = (k^2 - H_l)|kl\downarrow\rangle = 0. \tag{2.5}$$

Furthermore, any solution is a linear combination of $|kl\uparrow\rangle$ and $|kl\downarrow\rangle$. Therefore, using $G_l = G_{0l} + G_{0l} T_l G_{0l}$ we obtain,

$$|kql\uparrow\rangle = (1 + G_{0l}T_l)|ql\uparrow\rangle_0 + c_1|kl\uparrow\rangle + c_2|kl\downarrow\rangle. \tag{2.6}$$

 0022-2488/79/122520-06$01.00

In order to determine c_1 and c_2 we consider the asymptotic behavior of the right-hand side. By using

$$\langle r|G_l|r'\rangle = -\tfrac{1}{2}\pi k\,\langle r_>|kl\uparrow\rangle\langle kl-|r_<\rangle,$$

we obtain, for $r\to\infty$,

$$\langle r|G_{0l}\,T_l|ql\uparrow\rangle_0 = \langle r|G_l\,V_l|ql\uparrow\rangle_0$$

$$\sim -\tfrac{1}{2}\pi k\,\langle r|kl\uparrow\rangle\langle kl-|V_l|ql\uparrow\rangle_0, \qquad r\to\infty.$$

Since $\langle r|kql\uparrow\rangle$ has by definition the same asymptotic behavior as $\langle r|ql\uparrow\rangle_0$, namely,

$$\lim_{r\to\infty}\langle r|kql\uparrow\rangle qre^{-iqr} = (2/\pi)^{1/2},$$

we find

$$c_1 = \tfrac{1}{2}\pi k\,\langle kl-|V_l|ql\uparrow\rangle_0,$$
$$c_2 = 0.$$

It is convenient to rewrite c_1 in terms of the *off-shell Jost function* $f_l(k,q)$. Fuda[4] has given a closed expression which in our notation reads,

$$f_l(k,q) = 1 + \tfrac{1}{2}\pi q(q/k)^l f_l(k)\langle kl-|V_l|ql\uparrow\rangle_0. \tag{2.7a}$$

Some equivalent expressions are

$$f_l(k,q) = 1 + \tfrac{1}{2}\pi q(q/k)^l f_l(k)\langle kl|T_l|ql\uparrow\rangle_0, \tag{2.7b}$$

$$= 1 + \tfrac{1}{2}\pi q(q/k)^l f_l(k)\,{}_0\langle ql\downarrow|T_l|kl\rangle, \tag{2.7c}$$

$$= 1 + \tfrac{1}{2}\pi q(q/k)^l f_l(k)\,{}_0\langle ql\downarrow|V_l|kl+\rangle. \tag{2.7d}$$

By substituting c_1 in Eq. (2.6) and using (2.7a) we obtain the convenient expression,

$$|kql\uparrow\rangle = (1+G_{0l}\,T_l)|ql\uparrow\rangle_0 + |kl\uparrow\rangle(k/q)^{l+1}$$

$$\times\frac{f_l(k,q)-1}{f_l(k)}. \tag{2.8}$$

From now on we shall suppress the argument k of the Jost function, so we write f_l instead of $f_l(k)$.

When the potential has a short range the off-shell Jost function and solution are continuous in $q=k$, (cf. Ref. 3)

$$\lim_{q\to k} f_l(k,q) = f_l, \tag{2.9}$$

$$\lim_{q\to k}|kql\uparrow\rangle = |kl\uparrow\rangle. \tag{2.10}$$

By taking the limit $q\to k$ in Eq. (2.8) we obtain

$$|kl\uparrow\rangle = (1+G_{0l}\,T_l)|kl\uparrow\rangle_0 f_l. \tag{2.11}$$

We multiply both sides of this equation by V_l and get

$$V_l|kl\uparrow\rangle = T_l|kl\uparrow\rangle_0 f_l. \tag{2.12}$$

This equality turns out to be very useful below.

Multiplying Eq. (2.8) by V_l and using Eq. (2.12), we obtain

$$V_l|kql\uparrow\rangle = T_l|ql\uparrow\rangle_0 + T_l|kl\uparrow\rangle_0(k/q)^{l+1}$$

$$\times[f_l(k,q)-1]. \tag{2.13}$$

Further we get some closed formulas for the Jost function from Eqs. (7) by taking $q=k$. We have

$$f_l = f_l(k,k) = 1 + \tfrac{1}{2}\pi k f_l\langle kl-|V_l|kl\uparrow\rangle_0,$$

and therefore

$$f_l^{-1} = 1 - \tfrac{1}{2}\pi k\,\langle kl-|V_l|kl\uparrow\rangle_0$$

$$= 1 - \tfrac{1}{2}\pi k\,\langle kl|T_l|kl\uparrow\rangle_0$$

$$= 1 - \tfrac{1}{2}\pi k\,{}_0\langle kl\downarrow|V_l|kl+\rangle$$

$$= 1 - \tfrac{1}{2}\pi k\,{}_0\langle kl\downarrow|T_l|kl\rangle. \tag{2.14}$$

By using Eq. (2.12) we obtain from Eq. (2.14),

$$f_l = 1 + \tfrac{1}{2}\pi k\langle kl|V_l|kl\uparrow\rangle = 1 + \tfrac{1}{2}\pi k\langle kl\downarrow|V_l|kl\rangle. \tag{2.15}$$

We shall need the connection between $\langle p|kl\uparrow\rangle$ and $\langle p|V|kl\uparrow\rangle$. From Eq. (2.11) we have

$$|kl\uparrow\rangle f_l^{-1} = |kl\uparrow\rangle_0 + G_{0l}T_l|kl\uparrow\rangle_0,$$

$$= |kl\uparrow\rangle_0 + G_{0l}\,V_l|kl\uparrow\rangle f_l^{-1}. \tag{2.16}$$

Therefore,

$$\langle p|kl\uparrow\rangle = \langle p|kl\uparrow\rangle_0 f_l + \langle p|G_{0l}\,V_l|kl\uparrow\rangle. \tag{2.17}$$

The free "state" $\langle p|kl\uparrow\rangle_0$ is given explicitly by

$$\langle p|kl\uparrow\rangle_0 = \frac{2}{\pi k}\frac{(p/k)^l}{p^2-k^2}, \quad \mathrm{Im}k>0. \tag{2.18}$$

By inserting this in Eq. (2.17) one easily obtains

$$\langle p|V_l|kl\uparrow\rangle = (k^2-p^2)\langle p|kl\uparrow\rangle + 2(\pi k)^{-1}(p/k)^l f_l, \tag{2.19}$$

which is the relation we wanted.

The connection between the *off-shell* quantities, corresponding to the one of Eq. (2.19), can be obtained from Eqs. (2.8), (2.13), and (2.19),

$$\langle p|V_l|kql\uparrow\rangle = (k^2-p^2)\langle p|kql\uparrow\rangle$$
$$+ 2(\pi q)^{-1}(p/q)^l[f_l(k,q)$$
$$- (k^2-q^2)/(p^2-q^2)]. \tag{2.20}$$

It is interesting to consider the limit of $\langle p|kl\uparrow\rangle$ for $p\to\infty$. This limit could be used for an alternative definition of f_l (cf. Refs. 5 and 6). By using the fact that $\langle r|kl\uparrow\rangle = O(r^{-l-1})$ as $r\to 0$, we obtain

$$\langle p|V_l|kl\uparrow\rangle = (2/\pi)^{1/2}i^{-l}\int_0^\infty j_l(pr)V_l(r)\langle r|kl\uparrow\rangle r^2\,dr$$

$$\simeq cp^{l-2}\int_0^\infty j_l(x)V_l(x/p)x^{1-l}\,dx, \quad p\to\infty.$$

In this way we find that

$$\lim_{p\to\infty} p^{-l}\langle p|\,V_l\,|kl\uparrow\rangle = 0, \tag{2.21}$$

when the potential is nonsingular, i.e.,

$$V(r) = O(r^{-\alpha}), \quad \alpha < 2, \quad r\to 0.$$

It is easily seen from Eqs. (2.19) and (2.21) that

$$f_l = \tfrac{1}{2}\pi k^{l+1} \lim_{p\to\infty} p^{2-l}\langle p|\,kl\uparrow\rangle. \tag{2.22}$$

This may be compared with the usual definition of f_l,

$$f_l \equiv (\pi/2)^{1/2}(2/i)^l (l!/(2l)!)\lim_{r\to 0}(kr)^{l+1}\langle r|kl\uparrow\rangle. \tag{2.23}$$

Similar equations hold for the *off-shell* Jost function and solution. The analog of Eq. (2.23) is (Ref. 3)

$$f_l(k,q) = (\pi/2)^{1/2}(2/i)^l\,[l!/(2l)!]\lim_{r\to 0}(qr)^{l+1}\langle r|kql\uparrow\rangle. \tag{2.24}$$

The off-shell analog of Eq. (2.22) follows by using Eq. (2.8). We have [cf. Eq. (2.21)]

$$\lim_{p\to\infty} p^{2-l}\langle p|\,G_{0l}\,T_l\,|ql\uparrow\rangle_0 = \lim_{p\to\infty} p^{-l}\langle p|\,T_l\,|ql\uparrow\rangle_0 = 0,$$

and so we obtain from (2.8),

$$f_l(k,q) = \tfrac{1}{2}\pi q^{l+1}\lim_{p\to\infty} p^{2-l}\langle p|kql\uparrow\rangle. \tag{2.25}$$

This expression can also be derived with the help of Eq. (2.20).

It is interesting to note that Eq. (2.25) is obtained in a different way, by using Eq. (2.24) in the expression

$$\langle p|kql\uparrow\rangle = (2/\pi)^{1/2} i^{-l}\int_0^\infty j_l(pr)\langle r|kql\uparrow\rangle r^2\,dr, \tag{2.26}$$

and applying the equality

$$\int_0^\infty j_l(x)x^{1-\lambda}\,dx = \pi^{1/2}2^{-\lambda}\frac{\Gamma(1+\frac{1}{2}l-\frac{1}{2}\lambda)}{\Gamma(\frac{1}{2}+\frac{1}{2}l+\frac{1}{2}\lambda)},$$
$$0 < \mathrm{Re}\lambda < l+2. \tag{2.27}$$

On the other hand, we shall now derive Eq. (2.24) from Eq. (2.25). We have

$$\langle r|kql\uparrow\rangle = (2/\pi)^{1/2}\, i^l \lim_{\epsilon\downarrow 0}\int_0^\infty j_l(pr)\langle p|kql\uparrow\rangle e^{-\epsilon p}p^2\,dp, \tag{2.28}$$

where $e^{-\epsilon p}$ has been inserted to guarantee the convergence of the integral. It turns out that, when r goes to zero, $\langle p|kql\uparrow\rangle$ may be replaced by its asymptotic value, which is given by Eq. (2.25). Then we obtain from (2.28), using the new variable of integration $x = pr$,

$$\lim_{r\to 0}(qr)^{l+1}\langle r|kql\uparrow\rangle = f_l(k,q)(2/\pi)^{3/2} i^l \times \lim_{\epsilon\downarrow 0}\int_0^\infty j_l(x)x^l e^{-\epsilon x/r}\,dx. \tag{2.29}$$

In order to evaluate the integral here, we note that

$$\int_0^\infty e^{-\alpha x}J_\mu(\beta x)x^\nu\,dx = (\alpha^2+\beta^2)^{-(1/2)-(1/2)\nu} \times \Gamma(1+\mu+\nu)P_\nu^{-\mu}(\alpha(\alpha^2+\beta^2)^{-(1/2)}),$$
$$\alpha > 0, \quad \beta > 0, \quad \mathrm{Re}(1+\mu+\nu) > 0. \tag{2.30}$$

Here $P_\nu^{-\mu}(\zeta)$ is the Legendre function of the first kind "on the cut": $-1 < \zeta < 1$. Its value for $\zeta = 0$ is given by

$$\Gamma(1+\mu)P_\nu^{-\mu}(0) = {}_2F_1(-\nu,\nu+1;\mu+1;\tfrac{1}{2}) = \frac{\Gamma(1+\frac{1}{2}\mu)\Gamma(\frac{1}{2}+\frac{1}{2}\mu)}{\Gamma(1+\frac{1}{2}\mu+\frac{1}{2}\nu)\Gamma(\frac{1}{2}+\frac{1}{2}\mu-\frac{1}{2}\nu)}. \tag{2.31}$$

By using this expression we get

$$\lim_{\epsilon\downarrow 0}\int_0^\infty e^{-\epsilon x}J_\mu(x)x^\nu\,dx = 2^\nu\frac{\Gamma(\frac{1}{2}+\frac{1}{2}\mu+\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}+\frac{1}{2}\mu-\frac{1}{2}\nu)},$$
$$\mathrm{Re}(1+\mu+\nu) > 0,$$

and so

$$\lim_{\epsilon\downarrow 0}\int_0^\infty e^{-\epsilon x}j_l(x)x^l\,dx = \pi^{1/2}2^{l-1}\Gamma(l+\tfrac{1}{2}) = \pi 2^{-l-1}(2l)!/l!. \tag{2.32}$$

By inserting this in Eq. (2.29) we just obtain Eq. (2.24).

We note that the above limiting procedures constitute in fact a generalization of the well-known Riemann–Lebesgue lemma, i.e.,

$$\lim_{y\to\pm\infty}\int f(x)e^{ixy}\,dx = 0,$$

where f is any summable function.

3. THE COULOMB CASE

Some of the expressions derived in Sec. 2 do not hold when the potential has a Coulomb range. Especially Eqs. (2.9), (2.10), (2.14), and (2.15) need modification. In this section we shall derive the analogs of these equations for the case of the pure Coulomb potential. Further we shall develop some explicit expressions, in terms of hypergeometric functions, for the particular case when $l = 0$. In Sec. 4 we shall derive interesting formulas for the case when the potential is the sum of the Coulomb potential and a short-range potential, by using the results obtained in Secs. 2 and 3.

In the first place we note that the important equations (2.11) and (2.12) *do* hold for the Coulomb case, i.e.,

$$|kl\uparrow\rangle_c = (1+G_{0l}\,T_{cl})|kl\uparrow\rangle_0 f_{cl}, \tag{3.1}$$

and so

$$V_{cl}|kl\uparrow\rangle_c = T_{cl}|kl\uparrow\rangle_0 f_{cl}. \tag{3.2}$$

We shall prove Eq. (3.1) in an independent way. To start with, we observe that the *existence* of the quantity

$G_{0l}\,T_{cl}|kl\uparrow\rangle_0 = G_{cl}\,V_{cl}|kl\uparrow\rangle_0$ is easily confirmed by using

$$\langle r|\,G_{cl}\,|r'\rangle = -\tfrac{1}{2}\pi k\,\langle r_>|kl\uparrow\rangle_{cc}\langle kl-|r_<\rangle. \tag{3.3}$$

One can also show in this way that $G_{0l}T_{cl}|kl\downarrow\rangle_0$ is *not* defined, i.e., that it contains a divergent integral.

In order to prove Eq. (3.1), we note that its right-hand side equals some solution ψ of the equation $(k^2 - H_{cl})\psi = 0$ (in the coordinate representation). Further, by using Eq. (3.3) and by considering the asymptotic behavior ($r\to\infty$) of the right-hand side of (3.1), we find that it must be proportional to $|kl\uparrow\rangle_c$. The next step is to consider the behavior for small r. By again using (3.3) one has

$$\langle r|G_{cl}\,V_{cl}|kl\uparrow\rangle_0 = O(\ln r), \quad l=0, \quad r\to 0,$$

$$= O(r^{-l}), \quad l>0, \quad r\to 0. \tag{3.4}$$

Therefore,

$$\lim_{r\to 0} r^{l+1}\langle r|\,G_{0l}\,T_{cl}\,|kl\uparrow\rangle_0 = 0, \quad l=0,1,\cdots.$$

By using Eq. (2.23) the proof of Eq. (3.1) now follows easily.

In a previous paper[2] we have derived the Coulomb analog of Eq. (2.9), by using an explicit expression for $f_{cl}(k,q)$. The following equality holds,

$$\lim_{q\to k}\omega f_{cl}(k,q) = f_{cl}, \quad k>0. \tag{3.5}$$

Here

$$\omega \equiv \left(\frac{q-k}{q+k}\right)^{i\gamma}\frac{e^{\pi\gamma/2}}{\Gamma(1+i\gamma)} = \frac{f_{c0}}{f_{c0}(k,q)}, \quad \mathrm{Im}q>0. \tag{3.6}$$

The Coulomb analog of Eq. (2.10) is now easily obtained by using Eqs. (2.8), (3.1), and (3.5),

$$\lim_{q\to k}\omega|kql\uparrow\rangle_c = |kl\uparrow\rangle_c, \quad k>0. \tag{3.7}$$

It is interesting to note that we are now able to prove the validity of two conjectures from a preceding paper.[1] The first one, Eq. (40.l), is in fact just (3.7). The second one, Eq. (40.k), is easily proved by using Eqs. (40.h)–(40.j) of Ref. 1 and Eq. (2.8).

We note that Eqs. (2.7a) and (2.7d) are valid in the Coulomb case, whereas Eqs. (2.7b) and (2.7c) are not. By using Eqs. (2.7a) and (3.5) we have obtained the interesting equality,

$$f_{cl}^{-1} = \lim_{q\to k}(\omega^{-1} - \tfrac{1}{2}\pi k\,{}_c\langle kl-|\,V_{cl}\,|ql\uparrow\rangle_0), \quad k>0. \tag{3.8}$$

Obviously, this can be considered as a Coulomb analog of the short-range potential formulas given in Eq. (2.14).

It would be interesting to have available explicit expressions for the above quantities. Only a few such formulas are known. The quantities $\langle r|kl\uparrow\rangle_c$ and $f_{cl}(k)$ for $l=0,1,2,\cdots$ have been known for a long time. We have obtained a number of interesting analytic expressions for $f_{cl}(k,q)$, $l=0,1,2,\cdots$ (see Ref. 7). Further, $\langle p|T_{cl}|p'\rangle$ is known in closed form for $l=0$[8] and for $l=1$ only.[9] Below we shall derive analytic expressions for $\langle p|kl\uparrow\rangle_c$, $\langle p|kql\uparrow\rangle_c$, and for $\langle p\,|\,T_{cl}\,|ql\uparrow\rangle_0$, in the case $l=0$ only.

Before we start with the derivations we would like to make some remarks on the interrelationships between the above quantities. It is important to note that $\langle p|T_{cl}\,|ql\uparrow\rangle_0$ can be considered as the general object from which all other quantities can be obtained in a *simple* way. This is true as well for Coulomblike potentials, and of course also for short-range potentials. Indeed, by taking $q=k$ we have $\langle p|T_{cl}\,|kl\uparrow\rangle_0$ from which $\langle p|kl\uparrow\rangle_c$ follows with the help of Eqs. (2.12) and (2.19). Once $\langle p|kl\uparrow\rangle_c$ is known, $\langle p|kql\uparrow\rangle_c$ is obtained by using Eq. (2.8). The ordinary off-shell Coulomb T matrix $\langle p|T_{cl}\,|p'\rangle$ follows from $\langle\,p\,|\,T_{cl}\,|ql\uparrow\rangle_0$ by noting that

$$2iT_{cl}\,|p'l\,\rangle = T_{cl}\,|p'l\uparrow\rangle_0 + (-)^l T_{cl}\,|(-p')l\uparrow\rangle_0. \tag{3.9}$$

Furthermore, $f_{cl}(k,q)$ can be obtained from $\langle p|kql\uparrow\rangle_c$ and f_{cl} from $\langle p|kl\uparrow\rangle_c$ by using Eqs. (2.25) and (2.22), respectively. Finally we note that application of the Coulombian asymptotic states (see Ref. 10) to $\langle p|T_{cl}\,|ql\uparrow\rangle_0$ and $\langle p|T_{cl}\,|p'\rangle$ yields $f_{cl}(k,q)$ and $\langle p|kl+\,\rangle_c$, respectively. Since, therefore, $\langle p|T_{cl}\,|ql\uparrow\rangle_0$ appears to be the object of central important, we are interested in the general structure of an analytic expression for this quantity.

For the moment we restrict ourselves to the case $l=0$ and we suppress l. Let us first recall the expression for $\langle p|T_c|p'\rangle$ given in Ref. 8,

$$\langle p\,|\,T_c|p'\rangle = ik\,(\pi pp')^{-1}[F_{i\gamma}(aa') + F_{i\gamma}((aa')^{-1})$$

$$- F_{i\gamma}(a/a') - F_{i\gamma}(a'/a)]. \tag{3.10}$$

Here

$$F_{i\gamma}(z) \equiv {}_2F_1(1,i\gamma;\,1+i\gamma;\,z)$$

and

$$a \equiv \frac{(p-k)}{(p+k)}, \quad a' \equiv \frac{(p'-k)}{(p'+k)}.$$

By using a well-known integral representation for the hypergeometric function,

$$F_{i\gamma}(z) = i\gamma\int_0^1 t^{i\gamma-1}(1-tz)^{-1}\,dt, \tag{3.11}$$

we are able to evaluate

$$\langle p\,|\,T_c\,|q\uparrow\rangle_0 = \int_0^\infty \langle\,p\,|\,T_c\,|p'\rangle\langle p'|q\uparrow\rangle_0 p'^2\,dp',$$

where [cf. Eq. (2.18)]

$$\langle p|q\uparrow\rangle_0 = 2(\pi q)^{-1}(p^2-q^2)^{-1}, \quad \mathrm{Im}q>0.$$

After a number of manipulations we arrive at

$$\langle p\,|\,T_c|\,q\uparrow\rangle_0 = -2k\,(\pi pq)^{-1}[F_{i\gamma}(ab) - F_{i\gamma}(b/a)$$

$$- F_{i\gamma}(a) + F_{i\gamma}(1/a)], \tag{3.12}$$

with

$$b \equiv \frac{(q-k)}{(q+k)}.$$

Equation (3.9) provides us with a check on this result. It can be seen by inspection that we have indeed

$$\lim_{q\to p'}(\langle p|T_c|q\uparrow\rangle_0 + \langle p|T_c|(-q)\uparrow\rangle_0) = 2i\langle p|T_c|p'\rangle, \tag{3.13}$$

note that $b\to 1/b$ when q is replaced by $-q$.

Further, we clearly have from (3.12),

$$\lim_{q\to k}\langle p|T_c|q\uparrow\rangle_0 = \langle p|T_c|k\uparrow\rangle_0 = 2(\pi p)^{-1}[F_{i\gamma}(a) - F_{i\gamma}(1/a)]. \tag{3.14}$$

By using Eq. (3.1) one easily obtains

$$\langle p|V_c|k\uparrow\rangle_c = 2(\pi p)^{-1}f_c[F_{i\gamma}(a) - F_{i\gamma}(1/a)], \tag{3.15}$$

and with the help of Eq. (2.19),

$$\langle p|k\uparrow\rangle_c = 2(\pi p)^{-1}(p^2 - k^2)^{-1}f_c[p/k - F_{i\gamma}(a) + F_{i\gamma}(1/a)]. \tag{3.16}$$

Finally we note that $\langle p|kq\uparrow\rangle_c$ can now easily be given. We only need to insert the known expressions for the terms on the right-hand side of Eq. (2.8). In particular, we have

$$f_c(k,q) = b^{-i\gamma}. \tag{3.17}$$

Let us, for completeness, write out this expression for the Hankel transform of the Coulomb off-shell Jost solution for $l=0$,

$$\langle p|kq\uparrow\rangle_c = \frac{2}{\pi q(p^2-q^2)} + \frac{2k}{\pi pq(p^2-k^2)} \times [F_{i\gamma}(ab) - F_{i\gamma}(b/a) - p/k + b^{-i\gamma} \times \{p/k - F_{i\gamma}(a) + F_{i\gamma}(1/a)\}]. \tag{3.18}$$

By using Eq. (2.13) or Eq. (2.20) we have

$$\langle p|V_c|kq\uparrow\rangle_c = 2k(\pi pq)^{-1}[F_{i\gamma}(b/a) - F_{i\gamma}(ab) + b^{-i\gamma}\{F_{i\gamma}(a) - F_{i\gamma}(1/a)\}]. \tag{3.19}$$

By taking here the limit $q\to k$ we get, with $\omega = f_c b^{i\gamma}$,

$$\lim_{q\to k}\omega\langle p|V_c|kq\uparrow\rangle_c = \langle p|V_c|k\uparrow\rangle_c.$$

Such a relation holds in fact for all l. Indeed, with the help of Eqs. (2.19), (2.20), (3.5), and (3.7) the proof of

$$\lim_{q\to k} f_c b^{i\gamma}\langle p|V_{cl}|kql\uparrow\rangle_c = \langle p|V_{cl}|kl\uparrow\rangle_c, \quad l=0,1,2,\cdots, \tag{3.20}$$

is easily obtained.

A final remark concerning the generalization of the $l=0$ expression for $\langle p|T_c|q\uparrow\rangle_0$ to general values of l is appropriate here. In view of Eq. (3.12) it can be expected that $\langle p|T_{cl}|ql\uparrow\rangle_0$ where $l=0,1,2,\cdots$, can be expressed in terms of simple functions and the hypergeometric function $F_{i\gamma}$ with *exactly* the same arguments as in (3.12), notably ab, b/a, a, and $1/a$.

4. THE COULOMBLIKE POTENTIAL CASE

In this section we assume that the potential is the sum of the Coulomb potential and a short-range potential, $V = V_c + V_s$. We shall discuss the necessary modifications of the equations given in Sec. 2 by using the appropriate results obtained in Sec. 3. In particular, we will derive the analogs of Eqs. (2.9), (2.10), and (2.14).

We shall use the well-known two-potential formalism. The T operator corresponding to $V = V_c + V_s$ is given by

$$T = T_c + (1 + T_c G_0)t_{cs}(1 + G_0 T_c), \tag{4.1}$$

where t_{cs} is the solution of

$$t_{cs} = V_s + V_s G_c t_{cs}. \tag{4.2}$$

The partial-wave analogs of these equations have exactly the same form. For the partial-wave "outgoing" scattering state $|kl+\rangle$ the following equation can be obtained,

$$|kl+\rangle = |kl+\rangle_c + G_{cl}\, t_{csl}|kl+\rangle_c. \tag{4.3}$$

In order to derive relations for the "Jost states," we use Eqs. (2.11) and (2.12). These are also valid for a Coulomblike potential. We insert (4.1) in (2.11),

$$|kl\uparrow\rangle f_l^{-1} = (1 + G_{0l}\, T_l)|kl\uparrow\rangle_0,$$

and obtain

$$|kl\uparrow\rangle f_l^{-1} = |kl\uparrow\rangle_c f_{cl}^{-1} + G_{cl}\, t_{csl}|kl\uparrow\rangle_c f_{cl}^{-1} = (1 + G_{cl}\, t_{csl})|kl\uparrow\rangle_c f_{cl}^{-1}. \tag{4.4}$$

Further, by inserting (4.1) in (2.12),
$V_l|kl\uparrow\rangle f_l^{-1} = T_l|kl\uparrow\rangle_0$,
we get

$$V_l|kl\uparrow\rangle f_l^{-1} = V_{cl}|kl\uparrow\rangle_c f_{cl}^{-1} + G_{0l}^{-1} G_{cl}\, t_{csl}|kl\uparrow\rangle_c f_{cl}^{-1}. \tag{4.5}$$

We are now going to derive a connection between the Jost function f_l and the Coulomb Jost function f_{cl}. To this end we write Eq. (4.3) in the coordinate representation. In the resulting equation we insert the equality [cf. Eq. (3.3)]

$$\langle r|G_{cl}|r'\rangle = -\tfrac{1}{2}\pi k\langle r_<|kl+\rangle_c\, {}_c\langle kl\downarrow|r_>\rangle.$$

We note that ${}_c\langle kl\downarrow|t_{csl}|kl+\rangle_c$ is a well-defined quantity since t_{csl} is a short-range operator. By using

$$\langle r|t_{csl}|kl+\rangle_c = O(r^{l-\alpha}), \quad \alpha<2, \quad r\to 0,$$

we obtain from Eq. (4.3)

$$\langle r|kl+\rangle = \langle r|kl+\rangle_c - \tfrac{1}{2}\pi k\langle r|kl+\rangle_c \times {}_c\langle kl\downarrow|t_{csl}|kl+\rangle_c + O(r^{l+2-\alpha}), \quad r\to 0.$$

The Jost functions can be obtained from the scattering states

by considering their small-r behavior. We have (e.g., Refs. 5 and 6)

$$\lim_{r\to 0} r^{-l}\langle r|kl+\rangle = f_l^{-1}(2/\pi)^{1/2}(2ik)^l l!/(2l+1)!. \tag{4.6}$$

With the help of this relation one obtains

$$f_l^{-1} = f_{cl}^{-1} - \tfrac{1}{2}\pi k f_{cl}^{-1}\ {}_c\langle kl\downarrow|t_{csl}|kl+\rangle_c,$$

as can be seen by inspection. We rewrite this equation in the more convenient form,

$$f_{cl}f_l^{-1} = 1 - \tfrac{1}{2}\pi k\ {}_c\langle kl\downarrow|t_{csl}|kl+\rangle_c. \tag{4.7a}$$

If we take here $V_c \to 0$ we get back one of the expressions of Eq. (2.14), since in this case $t_{csl}\to T_{sl}$ and $f_{cl}\to 1$. Just as in (2.14) there are three different equivalent expressions, namely

$$f_{cl}f_l^{-1} = 1 - \tfrac{1}{2}\pi k\ {}_c\langle kl\downarrow|V_{sl}|kl+\rangle, \tag{4.7b}$$

$$= 1 - \tfrac{1}{2}\pi k\langle kl-|V_{sl}|kl\uparrow\rangle_c \tag{4.7c}$$

$$= 1 - \tfrac{1}{2}\pi k\ {}_c\langle kl-|t_{csl}|kl\uparrow\rangle_c. \tag{4.7d}$$

These are easily derived with the help of Eqs. (4.1) and (4.3).

In order to derive the analog of Eq. (2.15), we first multiply both sides of Eq. (4.4) by V_{sl}. This yields

$$V_{sl}|kl\uparrow\rangle f_l^{-1} = t_{csl}|kl\uparrow\rangle_c f_{cl}^{-1}. \tag{4.8}$$

By inserting this equation in (4.7d) we get

$$f_{cl}^{-1}f_l = 1 + \tfrac{1}{2}\pi k\ {}_c\langle kl-|V_{sl}|kl\uparrow\rangle$$

$$= 1 + \tfrac{1}{2}\pi k\langle kl\downarrow|V_{sl}|kl+\rangle_c. \tag{4.9}$$

Obviously this is the two-potential analog of Eq. (2.15).

It is interesting to consider the analog of Eq. (3.8), i.e.,

$$f_l^{-1} = \lim_{q\to k}(\omega^{-1} - \tfrac{1}{2}\pi k\langle kl-|V_l|ql\uparrow\rangle_0), \quad k>0. \tag{4.10}$$

In order to prove this equation, we first note that one has from Eqs. (4.1)–(4.3),

$$\langle kl-|V_l = {}_c\langle kl-|V_{cl} + {}_c\langle kl-|t_{csl}(1+G_{0l}T_{cl}). \tag{4.11}$$

We insert this expression in (4.10) and use

$$\lim_{q\to k}(1+G_{0l}T_{cl})|ql\uparrow\rangle_0 = |kl\uparrow\rangle_c f_{cl}^{-1}. \tag{4.12}$$

By applying finally Eqs. (3.8) and (4.7d) the proof of Eq. (4.10) is completed.

Now we turn to the *off-shell* Jost function. In Eq. (2.7a) the following general formula,

$$f_l(k,q) = 1 + \tfrac{1}{2}\pi q(q/k)^l f_l\langle kl-|V_l|ql\uparrow\rangle_0, \tag{4.13}$$

has been given.[4] This equation is also valid for a Coulomblike potential. By inserting Eq. (4.11) in (4.13), and by using (4.13) for the pure Coulomb case, we obtain

$$[f_l(k,q)-1]f_l^{-1} = [f_{cl}(k,q)-1]f_{cl}^{-1} + \tfrac{1}{2}\pi q(q/k)^l \times {}_c\langle kl-|t_{csl}(1+G_{0l}T_{cl})|ql\uparrow\rangle_0. \tag{4.14}$$

Herewith we have obtained a useful relation between the Coulomb off-shell Jost function and the off-shell Jost function for a Coulomblike potential. Indeed, from Eq. (4.14) one obtains, by using Eqs. (4.12) and (3.5), the analog of the pure Coulomb formula (3.5),

$$\lim_{q\to k}\omega f_l(k,q) = f_l, \quad k>0. \tag{4.15}$$

Here ω is given by Eq. (3.6).

Finally, we are going to prove

$$\lim_{q\to k}\omega|kql\uparrow\rangle = |kl\uparrow\rangle, \quad k>0. \tag{4.16}$$

This is just the Coulomblike analog of the pure Coulomb formula (3.7). From Eqs. (2.8) and (2.11) we obtain

$$|kql\uparrow\rangle \underset{q\to k}{\to} |kl\uparrow\rangle f_l^{-1}f_l(k,q).$$

Application of Eq. (4.15) then completes the proof of Eq. (4.16).

So we see that the singular behavior of the off-shell Jost function and of the off-shell Jost state in $q=k$ is just the same as for the pure Coulomb potential. This result is as might be expected, since this singularity is generated by the asymptotic part of the potential only.

ACKNOWLEDGMENT

This investigation forms a part of the research program of the Foundation for Fundamental Research of Matter (FOM), which is financially supported by the Netherlands Organization for Pure Scientific Research (ZWO).

[1] H. van Haeringen, Phys. Rev. A **18**, 56 (1978).
[2] H. van Haeringen, J. Math. Phys. **19**, 1379 (1978).
[3] M.G. Fuda and J.S. Whiting, Phys. Rev. C **8**, 1255 (1973).
[4] M.G. Fuda, Phys. Rev. C **14**, 37 (1976).
[5] J.R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[6] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[7] H. van Haeringen, J. Math. Phys. **20**, 1109 (1979).
[8] H. van Haeringen and R. van Wageningen, J. Math. Phys. **16**, 1441 (1975).
[9] H. van Haeringen, J. Math. Phys. **18**, 927 (1977).
[10] H. van Haeringen, J. Math. Phys. **17**, 995 (1976).

Symmetries of the stationary Einstein–Maxwell field equations. V[a)]

C. Hoenselaers

Department of Physics, Montana State University, Bozeman, Montana 59717
(Received 23 October 1978)

This paper shows that the field equations and the hierarchy of potentials for static electrovac fields can be formulated in close analogy to the stationary vacuum ones. A list of transformations, some of them previously unknown, will be given for the later case.

1. INTRODUCTION

In a series of papers[1] Kinnersley and Chitre explored the symmetry group of the stationary Einstein–Maxwell equations. In the present work, dealing with some aspects of the electrostatic Einstein–Maxwell equations and the stationary vacuum, we shall in Secs. 2 and 3 show that the equations for static electrovac space–times allow a treatment which differs only in the use of dual instead of complex functions from the one for vacuum. The generators of their symmetry group $\mathcal{K}$, a subgroup of the group discussed in II, will be identified with some of the generators of K'.

Section 4 deals with the stationary vacuum equations by the Lagrangian method as outlined in Refs. 2 and 3. We shall give a detailed prescription how to perform the Legendre transformations connecting the Lagrangians, from which the field equations are to be derived. We shall furthermore identify the generators of the invariance groups of those Lagrangians (at least for the first ten) with some of the $\gamma_{AB}^{(k)}$ used in Ref. 1.

While this method may be less powerful, it has the advantage that only a finite number of potentials are involved and the integration of the infinitesimal transformations is straightforward. Even though asymptotically flat metrics yield asymptotically nonflat ones, one may hope that some of the transformations can be used to cure the undesirable behavior of the later.

2. FIELD EQUATIONS AND THE HIERARCHY

Following Kinnersley,[4] we write the relevant equations for electrostatic space–times as

$$\frac{l}{\rho}\nabla(\rho\nabla l) = (\nabla l)^2 + (\nabla\psi)^2 \tag{2.1a}$$

$$\nabla\left(\frac{\rho}{l^2}\nabla\psi\right) = 0, \tag{2.1b}$$

where the Ernst potential $\mathscr{E}$ and the electromagnetic potential Ψ have been taken as real and we defined

$$l^2 = f.$$

It has been observed by Bonnor[5] that the above equations can be obtained from the vacuum Ernst equations by the substitution

$$\mathrm{Re}\mathscr{E} \to l^2, \quad \mathrm{Im}\mathscr{E} \to i\psi,$$

which, however, requires an analytic continuation. Thus it is not surprising that the symmetries of Eqs. (2.1) are very similar to the one of the stationary vacuum problem (cf., e.g., Refs. 2, 4, 6). The infinitesimal transformations, neglecting gauge transformations, read

$$l \to l - 2\eta l\psi, \quad \psi \to \psi - \eta(l^2 + \psi^2). \tag{2.2}$$

Equation (2.1b) implies the existence of a function ϕ defined by

$$\tilde{\nabla}\phi = \frac{\rho}{l^2}\nabla\psi. \tag{2.3}$$

Defining a quantity

$$l_{AB} = \begin{pmatrix} l & l\phi \\ l\phi & \rho^2 l^{-1} + l\phi^2 \end{pmatrix}, \tag{2.4}$$

Eqs. (2.1) can be written as

$$\nabla\left(\frac{1}{\rho} l_A{}^X \nabla l_{XB}\right) = 0, \tag{2.5}$$

or equivalently ($l^{XA} l_{XB} = \rho^2 \delta_B^A$)

$$\nabla l_{AB} = -\frac{1}{\rho} l_A{}^X \tilde{\nabla}\psi_{XB}, \quad \nabla\psi_{AB} = -\frac{1}{\rho} l_A{}^X \tilde{\nabla} l_{XB}. \tag{2.6}$$

To write those equations in a more concise way we make use of dual numbers[7] and define

$$h_{AB} = l_{AB} + j\psi_{AB},$$

which gives the field equation the desired form

$$\nabla h_{AB} = -\frac{j}{\rho} l_A{}^X \tilde{\nabla} h_{XB}. \tag{2.7}$$

This form of the field equations is, save for the use of dual instead of complex numbers, the same as (II 1.19). We hence can, with only minor changes in sign, define a hierarchy of fields and potentials in complete analogy to II:

$$\nabla h_{AB}^{(n)} = -\frac{j}{\rho} l_A{}^X \tilde{\nabla} h_{XB}^{(n)}, \tag{2.8}$$

$$h_{AB}^{(n+1)} = j(n_{AB}^{(1n)} + h_{AX} h^{(n)X}{}_B), \tag{2.9}$$

$$\nabla n_{AB}^{(mn)} + \bar{h}_{XA}^{(m)} \nabla h^{(n)X}{}_A. \tag{2.10}$$

We define

$$h_{AB}^{(0)} = -j\epsilon_{AB},$$

[a)]Supported by National Science Foundation Grant PHY76-12246.

 0022-2488/79/122526-04$01.00

which leads to

$$n_{AB}^{(0n)} = jh_{AB}^{(n)},$$

and the above equations hold also for $n = 0$. Furthermore we have the relations

$$n_{AB}^{(m,n)} - \bar{n}_{AB}^{(m,n)} = \bar{h}_{XA}^{(m)} h^{(n)X}{}_B,$$
$$n_{AB}^{(m,n+k)} - n_{AB}^{(m+k,n)} = \sum_s n_{AX}^{(m,s)} n^{(k-s,n)X}{}_B. \tag{2.11}$$

3. THE GROUP $\mathscr{K}$

The symmetry group of (2.6) consists of real 2×2 matrices with unit determinant, i.e., it is isomorphic to SL(2,R) (cf. I). Note that, in contrary to the group G of the stationary vacuum, the above group gives rise to nontrivial new solutions.

The symmetry group of (2.1) is also isomorphic to SL(2,R); however, neglecting gauge transformations, the covariant generalization (cf. II) of (2.2) can be written as

$$l_{AB} \to l_{AB} - 2(\eta_{XY} l^{XY} \psi_{(AB)} - \eta_{AB} l^{XY} \psi_{XY}),$$

from which one obtains the action on h_{AB}:

$$h_{AB} \to h_{AB} - j(\eta^{XY} h_{AX} h_{YB} + \eta_A{}^Y n_{YB}^{(11)}) - \eta^X{}_B h_{AX}^{(2)}.$$

As in II one constructs the action of the infinitesimal transformations on the potentials $n_{AB}^{(m,n)}$ and finds

$$n_{AB}^{(m,n)} \to n_{AB}^{(m,n)} + \eta_{AY}^{(k)} n^{(m+k,n)Y}{}_B + \eta_{YB}^{(k)} n^{(m,n+k)}{}_A{}^Y - \eta^{(k)XY} \sum_s n_{AX}^{(m,s)} n_{YB}^{(k-s,n)}. \tag{3.1}$$

The infinitesimal parameters $\eta_{AB}^{(k)}$ have the form

$$\eta_{AB}^{(k)} = \eta_{(AB)}^{(k)} + j\nu\epsilon_{AB}^{(k)} \tag{3.2}$$

with a real symmetric part.

Our group $\mathscr{K}$ is a subgroup of K'. To identify the parameters $\eta_{AB}^{(k)}$ with $\gamma_{AB}^{(k)}, c_A^{(k)}$ of II, we apply the transformations of $\mathscr{K}$ and K' to flat space.

The $h_{AB}^{(n)}$ can easily be shown to be given by

$$h_{11}^{(n)} = (2r)^{n-1} P_{n-1}(\cos\vartheta),$$
$$h_{12}^{(n)} = -j(2r)^n P_n(\cos\vartheta),$$
$$h_{21}^{(n)} = j(2r)^n \frac{1}{2n} \sin\vartheta P_{n-1}^1(\cos\vartheta),$$
$$h_{22}^{(n)} = -(2r)^{n+1} \frac{1}{2(n+1)} \sin\vartheta P_n^1(\cos\vartheta). \tag{3.3}$$

Furthermore, $n_{11}^{(n1)} = n_{21}^{(n1)} = 0$. Thus only one term of the sum survives if (3.1) is expanded for $n_{11}^{(01)}$.

We find

$$\eta_{11}^{(n)}: \quad h_{11} \to 1 - j\eta(2r)^{n+1} P_{n+1}(\cos\vartheta),$$
$$\eta_{12}^{(n)}: \quad h_{11} \to 1 - 2\eta(2r)^n P_n(\cos\vartheta),$$
$$\eta_{22}^{(n)}: \quad h_{11} \to 1 - j\eta(2r)^{n-1} P_{n-1}(\cos\vartheta), \tag{3.4}$$

while $\nu^{(k)}$ does not have any effect at all. The only transformation which preserves static metrics is $\gamma_{12}^{(k)}$. As

$$f = \mathrm{Re}(H_{11}) = \mathrm{Re}(h_{11})^2, \quad \psi = \mathrm{Re}(\phi_1) = \mathrm{Jm}(h_{11}), \tag{3.5}$$

we can immediately identify

$$\gamma_{12}^{(k)} \leftrightarrow 2\eta_{12}^{(k)}. \tag{3.6}$$

Moreover, we find from (II 3.2) and (III 5.19) that if the $c_A^{(n)}$ transformations are applied to flat space, the result is

$$c_A^{(n)}: \quad \phi_1 \to -ic(2r)^n P_n(\cos\vartheta),$$
$$c_2^{(n)}: \quad \phi_1 \to c(2r)^{n-1} P_{n-1}(\cos\vartheta), \tag{3.7}$$

and now

$$\mathrm{Im}\, c_1^{(k+1)} \longleftrightarrow \eta_{11}^{(k)}, \quad \mathrm{Re}\, c_2^{(k)} \longleftrightarrow \eta_{22}^{(k)}. \tag{3.8}$$

One checks from commutators (II 3.4) that

$(\gamma_{12}^{(k)}, \mathrm{Im}\, c_1^{(k+1)}), \mathrm{Re}\, c_2^{(k)})$ form indeed the subgroup of K' leaving the electrostatic vacuum invariant. Of course, one can perform a duality rotation $\Phi \to e^{i\alpha}\phi$.

We sketch the proof of the relations

$$B^{-1}\gamma_{12}^{(k)}B = 2\gamma_{12}^{(k)}, \quad B^{-1}\gamma_{22}^{(k)}B = c_2^{(k)},$$
$$B^{-1}\gamma_{11}^{(k)}B = -c_1^{(k+1)},$$

where $c_2^{(k)}$ is real, $c_1^{(k)}$ imaginary, and B denotes the Bonnor transformation. One first establishes by direct calculation from (II 3.1), (II 3.2)

$$B^{-1}\gamma_{12}^{(0)}B = 2\gamma_{12}^{(0)}, \quad B^{-1}\gamma_{22}^{(0)}B = c_2^{(0)}, \quad B^{-1}\gamma_{22}^{(1)}B = c_2^{(1)},$$

and uses $[B^{-1}\gamma^{(k)}B, B^{-1}\gamma^{(l)}B] = B^{-1}[\gamma^{(k)},\gamma^{(l)}]B$ and the commutation relations (II 3.4) to show the above relations for $k = 0,1$. Then one proceeds by induction. Negative k are dealt with analogously.

4. LAGRANGIANS FOR THE STATIONARY VACUUM CASE

It has been pointed out in A that the stationary vacuum Einstein equations in the presence of two commuting surface-forming Killing vectors can be derived from a series of Lagrangians

$$L^{(2n)} = 2\nabla k \cdot \nabla\rho - \frac{1}{2\rho}|\nabla f_{AB}|\ [\rho^2 = -\det(f_{AB})] \tag{4.1}$$

and

$$L^{(2n+1)} = 2\nabla\kappa\cdot\nabla\rho - \frac{\rho}{2f^2}(\nabla f^2 + \nabla\psi^2). \tag{4.2}$$

(We use those forms of the Lagrangians as they seem to be the most familiar ones, and changed the notation slightly to conform with I–IV.) Note that, while in (4.1) ρ is just an abreviation for a more complicated expression, it is an independent variable in (4.2). Variation of (4.2) leads to the Ernst equations.

The invariance group of the Lagrangians is generated by the Killing vectors, i.e., by

$$\xi_\alpha{}^a = \begin{pmatrix} 0 \\ 0 \\ 2f_{12} \\ f_{11} \end{pmatrix}, \quad \xi_\beta{}^a = \begin{pmatrix} 0 \\ f_{11} \\ -f_{22} \\ 0 \end{pmatrix}, \quad \xi_\gamma{}^a = \begin{pmatrix} 0 \\ 2f_{12} \\ 0 \\ f_{22} \end{pmatrix} \tag{4.3}$$

and

$$\xi_\eta{}^a = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \xi_\epsilon{}^a = \begin{pmatrix} 0 \\ 0 \\ f \\ \psi \end{pmatrix}, \quad \xi_\zeta{}^a = \begin{pmatrix} 0 \\ 0 \\ 2f\psi \\ \psi^2 - f^2 \end{pmatrix}, \tag{4.4}$$

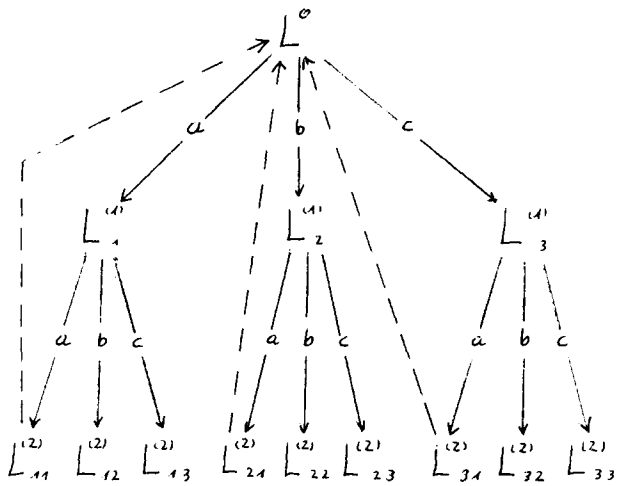

FIG. 1. The tree of Lagrangians. The letters on the solid lines refer to Eq. (4.7) and (4.10), respectively, while the dashed lines indicate that $\mathscr{G}^{(k+2)}{}_{\ldots ab1}$ is identical to $\mathscr{G}^{k}{}_{\ldots a}$.

where the sequence of variables is $(k, f_{11}, f_{22}, f_{12})$ and (κ, ρ, f, ψ), respectively.

The commutators of the transformations are the negative of the Lie bracket of the respective Killing vectors, i.e.,

$$[\beta,\alpha] = \alpha, \quad [\gamma,\alpha] = -2\beta, \quad [\gamma,\beta] = \gamma \tag{4.5}$$

and

$$[\epsilon,\eta] = -\eta, \quad [\zeta,\eta] = -2\epsilon, \quad [\zeta,\epsilon] = -\zeta, \tag{4.6}$$

which immediately shows that both groups are isomorphic to SL(2,R). There is also a Killing vector mapping $k \to k$ + const and $\kappa \to \kappa$ + const, respectively; we shall omit it.

Concentrating now on (4.1) and (4.3), we look for a hypersurface-orthogonal linear combination of the Killing vectors. A rather lengthy calculation shows that the equation

$$\left(\frac{1}{N}(\alpha\xi_{\alpha[a} + \beta\xi_{\beta[a} + \gamma\xi_{\gamma[a)}\right)_{,b]} = 0$$

with

$$N = \frac{1}{2\rho}[(2\gamma f_{12} + \beta f_{11})(2\alpha f_{12} - \beta f_{22}) - (\gamma f_{22} + \alpha f_{11})^2]$$

is satisfied for arbitrary α, β, γ. The Killing coordinate is, however, a rather complicated logarithmic function and leads us into the dead end street of Eq. (A8). We thus demand that N be a perfect square. This gives

$$\beta^2 = -4\alpha\gamma,$$

where we can resolve the following three cases: $\gamma = 0$, $-\alpha = \gamma = \frac{1}{2}\beta = 1$, $\alpha = 0$. We find the parametrizations

$$\begin{aligned} f_{11} &= f, \\ f_{12} &= -\omega f, \\ f_{22} &= \omega^2 f - \rho^2 f^{-1}, \\ \omega &= -f_{12}/f_{11}, \\ f_{11} &= \tfrac{1}{2}[f(1+\omega)^2 - \rho^2/f], \\ f_{22} &= \tfrac{1}{2}[f(1-\omega)^2 - \rho^2/f], \end{aligned} \tag{4.7a}$$

$$\begin{aligned} f_{12} &= \tfrac{1}{2}[f(1-\omega^2) + \rho^2/f], \\ \omega &= (f_{11} - f_{22})/(2f_{12} + f_{11} + f_{22}), \\ f &= \tfrac{1}{2}(f_{11} + f_{22}) + f_{12}, \end{aligned} \tag{4.7b}$$

$$\begin{aligned} f_{11} &= \omega^2 f - \rho^2 f^{-1}, \\ f_{12} &= -\omega f, \\ f_{22} &= f, \\ \omega &= -f_{12}/f_{22}. \end{aligned} \tag{4.7c}$$

(To avoid a profusion of primes, hats, or tildes here and in the following, we shall use them only when the danger of confusion is really imminent.)

The first and third of the above expressions have been known for a long time while the second one is new.

All those expressions yield with

$$\kappa = k + \tfrac{1}{2}\ln|f| \tag{4.8}$$

the Lagrangian

$$L^{(2n)} = 2\nabla\kappa\cdot\nabla\rho + \frac{1}{2}\left(\frac{f^2}{\rho}\nabla\omega^2 - \frac{\rho}{f^2}\nabla f^2\right),$$

from which a Legendre transformation with

$$\tilde{\nabla}\psi = \frac{f^2}{\rho}\nabla\omega \tag{4.9}$$

gives $L^{(2n+1)}$.

We now describe how to transform $L^{(2n+1)}$ into $L^{(2n+2)}$. One again searches for a hypersurface orthogonal linear combination of the Killing vector (4.4), and finds that any linear combination satisfies that requirement. One demands, for the same reasons as above, that the denominator be a square. The three cases to be distinguished are

$$g = f, \quad \chi = \psi, \tag{4.10a}$$

$$\begin{aligned} g &= \frac{2f}{f^2 + (\psi+1)^2}, \\ \chi &= \frac{1 - f^2 - \psi^2}{f^2 + (\psi+1)^2}, \\ f &= \frac{2g}{g^2 + (\chi+1)^2}, \\ \psi &= \frac{1 - g^2 - \chi^2}{g^2 + (\chi+1)^2}, \end{aligned} \tag{4.10b}$$

$$\begin{aligned} g &= \frac{f}{f^2 + \psi^2}, \\ \chi &= \frac{\psi}{f^2 + \psi^2}, \\ f &= \frac{g}{g^2 + \psi^2}, \\ \psi &= \frac{\chi}{g^2 + \chi^2}. \end{aligned} \tag{4.10c}$$

It should now be noted that the parametrization

$$\begin{aligned} g_{11} &= \frac{\rho}{f}, \\ g_{12} = g_{21} &= \frac{\rho\psi}{f}, \\ g_{22} &= \frac{\rho}{f}(f^2 + \psi^2), \\ \kappa' &= \kappa - \tfrac{1}{4}\ln\rho \end{aligned} \tag{4.11}$$

cast (4.2) in a form similar to (4.1) and the relations (4.10) turn out to be the analog of (4.7).

Equations (4.10) leave $L^{(2n+1)}$ form invariant, i.e., replace $f \to g$, $\psi \to \chi$, and, defining φ so that

$$\frac{\rho}{g^2}\tilde{\nabla}\chi = \nabla\varphi, \tag{4.12}$$

one gets $L^{(2n+2)}$. As the structure of the symmetry group of $L^{(2n+1)}$, being generated by the Killing vectors, is independent of the parametrization, we shall canonically choose (4.7a) for calculating the f_{AB} from g, φ which replaces f, ω.

5. THE GROUPS

In this section we shall identify the transformations generated by (4.3) and (4.4) with those generated by the $\gamma^{(k)}_{AB}$ of II and list them for $\mathcal{G}^{(1)}{}_a$ and $\mathcal{G}^{(2)}{}_{ab}$ (Fig. 1). ($\mathcal{G}^{(n)}{}_{ab\ldots}$ denotes the group belonging to $L^{(n)}{}_{ab\ldots}$).

One first observes that in general

$$\mathcal{G}^{(k+2)}{}_{\ldots ab\,1} = \mathcal{G}^{(k)}{}_{\ldots a}, \tag{5.1}$$

and that β and ϵ fulfill the relations

$$\begin{aligned}
&\beta(\mathcal{G}^{(2n)}{}_{\ldots a2}) = \tfrac{1}{2}(\zeta - \eta)(\mathcal{G}^{(2n-1)}{}_{\ldots a}),\\
&\beta(\mathcal{G}^{(2n)}{}_{\ldots a3}) = -\epsilon(\mathcal{G}^{(2n-1)}{}_{\ldots a}),\\
&\epsilon(\mathcal{G}^{(2n+1)}{}_{\ldots a2}) = \tfrac{1}{2}(\alpha + \gamma)(\mathcal{G}^{(2n)}{}_{\ldots a}),\\
&\epsilon(\mathcal{G}^{(2n+1)}{}_{\ldots a3}) = \beta(\mathcal{G}^{(2n)}{}_{\ldots a}),
\end{aligned} \tag{5.2}$$

respectively. Furthermore, α and $\eta(\mathcal{G}^{(n)})$ can contain only $\gamma^{(k)}_{AB}$ with $0 \leqslant k \leqslant -n$ and analogously for γ and ζ $(0 \leqslant k \leqslant n)$.

Now the commutators (4.5) and (4.6) have to be compared with the commutators for the $\gamma^{(k)}_{AB}$, i.e.,

$$[\gamma^{(k)}, \gamma^{(l)}] = \gamma^{(k+l)}, \quad \gamma^{(k+l)}_{AB} = 2\gamma^{(k)X}{}_{(A}\gamma^{(l)}_{BX)}$$

giving α, γ, respectively η, ζ, up to $\alpha \to \alpha\cdot\text{const}$, $\gamma \to \gamma\cdot(\text{const})^{-1}$. This remaining constant is then determined by the action on flat space and comparison with (III 6.1).

We find

$$\begin{aligned}
&\mathcal{G}^0[-\gamma^{(0)}_{22}, -\tfrac{1}{2}\gamma^{(0)}_{12}, \gamma^{(0)}_{11}],\\
&\mathcal{G}^{(1)}{}_1[-\tfrac{1}{2}\gamma^{(0)}_{12}, \gamma^{(-1)}_{11}, \gamma^{(1)}_{22}],\\
&\mathcal{G}^{(1)}{}_2[\tfrac{1}{2}(\gamma^{(0)}_{11} - \gamma^{(0)}_{22}), \tfrac{1}{2}(\gamma^{(-1)}_{11} + \gamma^{(-1)}_{12} + \gamma^{(-1)}_{22}),\\
&\quad \tfrac{1}{2}(\gamma^{(1)}_{11} - \gamma^{(1)}_{12} + \gamma^{(1)}_{22})],\\
&\mathcal{G}^{(1)}{}_3[\tfrac{1}{2}\gamma^{(0)}_{12}, -\gamma^{(-1)}_{22}, -\gamma^{(1)}_{11}],\\
&\mathcal{G}^{(2)}{}_{12}[\tfrac{1}{2}(\gamma^{(0)}_{22} + \gamma^{(-1)}_{12} + \gamma^{(-2)}_{11}), \tfrac{1}{2}(\gamma^{(1)}_{22} - \gamma^{(-1)}_{11}),\\
&\quad \tfrac{1}{2}(-\gamma^{(0)}_{11} + \gamma^{(1)}_{12} - \gamma^{(2)}_{22})],\\
&\mathcal{G}^{(2)}{}_{13}[\gamma^{(-2)}_{11}, \tfrac{1}{2}\gamma^{(0)}_{12}, -\gamma^{(2)}_{22}],\\
&\mathcal{G}^{(2)}{}_{22}[\tfrac{1}{4}(\gamma^{(0)}_{11} - \gamma^{(0)}_{12} + \gamma^{(0)}_{22} - 2\gamma^{(-1)}_{11} + 2\gamma^{(-1)}_{22}\\
&\quad + \gamma^{(-2)}_{11} + \gamma^{(-2)}_{12} + \gamma^{(-2)}_{22}),\\
&\quad \tfrac{1}{4}(\gamma^{(1)}_{11} - \gamma^{(1)}_{12} + \gamma^{(1)}_{22} - \gamma^{(-1)}_{11} - \gamma^{(-1)}_{12} - \gamma^{(-1)}_{22}),\\
&\quad \tfrac{1}{4}(\gamma^{(0)}_{11} + \gamma^{(0)}_{12} + \gamma^{(0)}_{22} + 2\gamma^{(1)}_{11} - 2\gamma^{(1)}_{22} + \gamma^{(2)}_{11} - \gamma^{(2)}_{12}\\
&\quad + \gamma^{(2)}_{22})],\\
&\mathcal{G}^{(2)}{}_{23}[-\tfrac{1}{2}(\gamma^{(-1)}_{11} - \gamma^{(-1)}_{12} + \gamma^{(-1)}_{22}), \tfrac{1}{2}(\gamma^{(0)}_{22} - \gamma^{(0)}_{11}),\\
&\quad -\tfrac{1}{2}(\gamma^{(1)}_{11} + \gamma^{(1)}_{12} + \gamma^{(1)}_{22})],\\
&\mathcal{G}^{(2)}{}_{32}[-\tfrac{1}{2}(\gamma^{(0)}_{11} + \gamma^{(-1)}_{12} + \gamma^{(-2)}_{22}), \tfrac{1}{2}(\gamma^{(-1)}_{22} - \gamma^{(1)}_{11}),\\
&\quad \tfrac{1}{2}(-\gamma^{(0)}_{22} + \gamma^{(1)}_{12} - \gamma^{(2)}_{11})],\\
&\mathcal{G}^{(2)}{}_{33}[\gamma^{(-2)}_{22}, -\tfrac{1}{2}\gamma^{(0)}_{12}, -\gamma^{(2)}_{11}].
\end{aligned}$$

The groups are given as $\mathcal{G}[\alpha, \beta, \gamma]$ and $\mathcal{G}[\epsilon, \eta, \zeta]$.

6. REMARKS

(i) $\mathcal{G}^0$, $\mathcal{G}^{(1)}{}_1$, and $\mathcal{G}^{(1)}{}_3$ have, of course, been known for a long time while $\mathcal{G}^{(1)}{}_2$ is new. The infinitesimal action $\zeta(\mathcal{G}^{(1)}{}_2)$ produces on flat space is

$$\begin{aligned}
&f_{11} \to 1 + 2\zeta z,\\
&f_{12} \to -2\zeta z\rho^2,\\
&f_{22} \to -\rho^2(1 - 2\zeta z).
\end{aligned} \tag{6.1}$$

Considering (4.7b) and (4.9), one finds that

$$\psi = \tfrac{1}{2}(\Omega_{11} + \Omega_{22}) + \Omega_{(12)} \tag{6.2}$$

in terms of the Ω_{AB} of II. These relations together with the structure of $\mathcal{G}^{(2)}{}_2$ shows that it is indeed one of the subgroups conjugate to H (cf. II).

(ii) It has been pointed out by Neugebauer and Kramer [3] that the mapping

$$f \to \rho/f, \quad \omega \to i\psi \tag{6.3}$$

maps (4.1), parametrized according to (4.7), directly into (4.2). Using (4.11) gives

$$L^{(2n+1)} = 2\nabla\kappa'\cdot\nabla\rho + \frac{1}{2\rho}|\nabla g_{AB}| \quad [\rho^2 = \det(g_{AB})]. \tag{6.4}$$

As, however, the field equations for the g_{AB} turn out to be

$$\nabla\left(\frac{1}{\rho}g_A{}^X\nabla g_{XB}\right) = 0, \tag{6.5}$$

it is clear that one can repeat the discussion of Secs. 2 and 3 almost word for word. The above transformation has the effect of, so to say, starting the infinite sequence of potentials at the first instead of the zeroth level.

Note added in proof: A change in the printing method prevents us from using the notation of I–IV. Overhead letters now appear as right superscripts in parentheses.

[1] W. Kinnersley, J. Math. Phys. **18**, 1529 (1977); W. Kinnersley and D.M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926, (1978); **19**, 2037 (1978); henceforth referred to as I–IV.

[2] C. Hoenselaers, J. Math. Phys. **17**, 1264 (1976). Lacking the Roman numeral "0", this paper is referred to as A.

[3] G. Neugebauer and D. Kramer, Ann. Phys. **24**, 62 (1969).

[4] W. Kinnersley, J. Math. Phys. **14**, 651 (1973).

[5] W. Bonnor, Z. Phys. **190**, 444 (1966).

[6] R. Geroch, J. Math. Phys. **12**, 918 (1971).

[7] Dual numbers are defined by $d := a + jb$ $(a, b \in \mathbb{R}, j^2 = 1)$. The dual conjugate is $\bar{d} = a - jb$ and the absolute value $|d| = (d\bar{d})^{1/2}$. They do not form a field as multiples of $1 + j$ or its conjugate do not have an inverse.

Symmetries of the stationary Einstein–Maxwell equations. VI. Transformations which generate asymptotically flat spacetimes with arbitrary multipole moments[a)]

C. Hoenselaers, William Kinnersley, and Basilis C. Xanthopoulos

Department of Physics, Montana State University, Bozeman, Montana 59717
(Received 24 January 1979)

A new series of transformations is presented for generating stationary axially symmetric asymptotically flat vacuum solutions of Einstein's equations. The application requires only algebraic manipulations to be performed. Several examples are given of new stationary axisymmetric solutions obtained in this way. It is conjectured that the transformations, applied to the general Weyl metric, can be used to generate systematically all stationary metrics with axial symmetry.

1. INTRODUCTION

In a continuing series of papers[1-5] (referred to as I–V) we have been studying the symmetry group **K** of the stationary, axially symmetric Einstein–Maxwell equations, and how the transformations $\gamma_{AB}^{(k)}$ may be used to generate new solutions of those equations. In Ref. 6 we found a set of transformations $\beta^{(k)}$ which automatically preserve asymptotic flatness. The $\beta^{(k)}$'s are finite linear combinations of the $\gamma_{AB}^{(k)}$'s, and have the further property that they leave flat space invariant. However, in practice those transformations are not easy to apply. The only successful examples found to date remain the generation of the Kerr and the generalized Tomimatsu–Sato metrics given in IV.

In the present paper we discuss a new set of transformations $A^{(p)}$ belonging to **K**, which seem to be much more interesting than the $\beta^{(k)}$'s for generating new solutions. By contrast with the $\beta^{(k)}$ the $A^{(p)}$ are built from infinite linear combinations of the $\gamma_{AB}^{(k)}$. They do not leave flat space invariant, but they map it into some asymptotically flat, nonflat spactime. However, their most interesting property is that they can be applied straightforwardly to any stationary, axially symmetric vacuum soluton. The procedure involves only algebraic manupulations and therefore can be easily adapted for use on a computer.

The general stationary axially symmetric asymptotically flat metric (if Newtonian gravity is any guide) should be characterized by two infinite sets of multipole moments, one for the mass distribution and one for the angular momentum distribution. The Weyl static metrics already contain the first set. Our belief is that the transformations $A^{(p)}$ generate the second set. Hence, we think that we have at hand an effective procedure for the construction of an arbitrary stationary, axially symmetric, asymptotically flat exact solution with any prescribed multipole moments.

2. GENERATING FUNCTIONS

In Paper IV the generating functions

$$F_{AB}(t) = \sum_{n=0}^{\infty} t^n H_{AB}^{(n)}, \tag{2.1}$$

$$G_{AB}(s,t) = \sum_{m,n=0}^{\infty} s^m t^n N_{AB}^{(m,n)}, \tag{2.2}$$

$$G_{AB}(0,t) = -iF_{AB}(t), \tag{2.3}$$

were introduced as an effective means for calculating the potentials $N_{AB}^{(m,n)}$. In the present paper we will find them to play an even more important role. The transformations we will presently consider can be described only in terms of the generating functions $G_{AB}(s,t)$. In this section, we will derive a few new properties of $G_{AB}(s,t)$ which will prove useful later. We will then present a calculation of $G_{AB}(s,t)$ for the general Weyl static metric.

$F_{AB}(t)$ obeys the differential equation

$$\nabla F_{AB} = itS^{-2}[(1-2tz)\nabla H_{AX} - 2t\rho\tilde{\nabla}H_{AX}]F^{X}{}_{B} \tag{2.4}$$

and the algebraic relation

$$2iftF_{2B} = (1 - 2tz - 2itf\omega)F_{1B} + SF_{1B}{}^{*}, \tag{2.5}$$

where

$$S(t) = [(1-2tz)^2 + (2t\rho)^2]^{1/2}. \tag{2.6}$$

[Geometrically $S(t)$ is $2t$ times the distance from the point $z = (2t)^{-1}, \rho = 0$.] $G_{AB}(s,t)$ is given in terms of $F_{AB}(t)$ by

$$G_{AB}(s,t) = (s-t)^{-1}[s\epsilon_{AB} - t(1-2sz)F_{XA}^{*}(s)F^{X}{}_{B}(t) - 2istf_{XY}F^{X}{}_{A}{}^{*}(s)F^{Y}{}_{B}(t)]. \tag{2.7}$$

These relationships are quoted as they appeared in IV, but they can be further simplified. From Eq. (2.5),

$$2itf_{AX}F^{X}{}_{B} = (1-2tz)F_{AB} + SF_{AB}^{*}, \tag{2.8}$$

and when this is used in Eq. (2.7), we obtain

$$G_{AB}(s,t) = (s-t)^{-1}[s\epsilon_{AB} + tS(s)F_{XA}(s)F^{X}{}_{B}(t)]. \tag{2.9}$$

Note that no pole is present at $s = t$, since the residue vanishes by virtue of

$$F_{XA}(s)F^{X}{}_{B}(s) = -S^{-1}(s)\epsilon_{AB}. \tag{2.10}$$

Two other functions of particular interest are

$$G'_{AB}(s) = \sum_{m=0}^{\infty} s^m N_{AB}^{(m,1)} = \left[\frac{\partial}{\partial t} G_{AB}(s,t)\right]_{t=0}$$

[a)]Partially supported by National Science Foundation Grant PHY 78-12294.

 0022-2488/79/122530-07$01.00

$$= s^{-1}[\epsilon_{AB} - iSF_{BA}], \tag{2.11}$$

$$G_{AB}(s,s) = \sum_{m,n=0}^{\infty} s^{m+n} N_{AB}^{(m,n)}$$

$$= \epsilon_{AB} + sSF^{X}{}_{A} \frac{d}{ds} F_{XB}. \tag{2.12}$$

(The last line was obtained using l'Hospital's rule.)

Potentials for the Weyl metric

The Weyl solutions are specified by the metric

$$f_{AB} = \begin{pmatrix} e^{2\chi} & 0 \\ 0 & -\rho^2 e^{-2\chi} \end{pmatrix}, \tag{2.13}$$

where $\nabla^2\chi = 0$. Then from Eqs. (II.1.12), (II.1.13),

$$\nabla H_{11} = 2e^{2\chi}\nabla\chi, \quad \nabla H_{12} = 2i\rho\widetilde{\nabla}\chi - i\widetilde{\nabla}\rho. \tag{2.14}$$

Put these in Eq. (2.4), with $A = B = 1$, and use Eq. (2.5) to eliminate F_{21}. The result is an equation for F_{11} alone

$$\nabla F_{11} = \{\nabla\chi - S^{-1}[\nabla S - (1-2tz)\nabla\chi + 2t\rho\widetilde{\nabla}\chi]\}F_{11}. \tag{2.15}$$

The solution is

$$F_{11}(t) = tS^{-1}(t)\exp[\chi + \beta(t)], \tag{2.16}$$

where $\beta(t)$ is a new potential, defined by

$$\nabla\beta = S^{-1}(t)[(1-2tz)\nabla\chi - 2t\rho\widetilde{\nabla}\chi]. \tag{2.17}$$

The other generating functions are now easily obtained:

$$F_{12}(t) = iS^{-1}\exp(\chi - \beta),$$

$$F_{21}(t) = -\tfrac{1}{2}iS^{-1}(1-2tz+S)\exp(\beta - \chi),$$

$$F_{22}(t) = \tfrac{1}{2}t^{-1}S^{-1}(1-2tz-S)\exp(-\beta - \chi),$$

$$G_{11}(s,t) = -\tfrac{1}{2}itS^{-1}(t)\exp[\beta(s) + \beta(t)] \times\left(1 + \frac{s+t-4stz}{sS(t) + tS(s)}\right), \tag{2.18}$$

$$G'_{11}(t) = -i\exp(\chi + \beta),$$

$$G_{11}(t,t) = F_{11}(t)F_{21}(t).$$

The potential $\beta(t)$ is a natural t-dependent generalization of the original harmonic function χ. One may show that $\beta(t)$ has the following properties:

$$\beta(0) = \chi, \quad [\nabla\beta(t)]^2 = [\nabla\chi]^2,$$

$$\nabla^2[\beta(t)/S(t)] = 0. \tag{2.19}$$

Equation (2.17), however, does not determine $\beta(t)$ uniquely. There remains the freedom

$$\beta(t) \to \beta(t) + C(t), \quad C(0) = 0, \tag{2.20}$$

where $C(t)$ is independent of the spatial coordinates ρ and z. This amounts to a gauge transformation, similar to the ones previously encountered in III, Sec. 5 and IV, Sec. 2.

One may further specify χ in terms of a multipole expansion

$$\chi = \sum_{n=0}^{\infty} m_n r^{-n-1} P_n(\cos\theta), \tag{2.21}$$

where r, θ are spherical coordinates related to Weyl coordinates by $z = r\cos\theta$, $\rho = r\sin\theta$, and the P_n are Legendre polynomials. We next seek the corresponding expression for $\beta(t)$. Using a series expansion for small t, we find the solution of Eq. (2.17) to be

$$\beta(t) = S(t)\sum_{n=0}^{\infty}\sum_{k=0}^{n} m_n (2t)^{n-k} r^{-k-1} P_k(\cos\theta). \tag{2.22}$$

On the other hand, for large t one finds

$$\beta(t) = \sum_{n=0}^{\infty} m_n[(2t)^{n+1} - r^{-n-1}P_{n+1}(\cos\theta) + O(t^{-1})]. \tag{2.23}$$

The divergent behavior at $t = \infty$ may be cured by making use of the gauge freedom. This gives the result

$$\beta(\infty) = -\sum_{n=0}^{\infty} m_n r^{-n-1} P_{n+1}(\cos\theta), \tag{2.24}$$

$$F_{11}(\infty) = \tfrac{1}{2}r^{-1}\exp[\chi + \beta(\infty)],$$

$$F_{21}(\infty) = -\tfrac{1}{2}i(1-\cos\theta)\exp[-\chi + \beta(\infty)], \tag{2.25}$$

$$F_{12}(\infty) = F_{22}(\infty) = 0.$$

Potentials for the Voorhees metrics

Finally we return briefly to consider the generating functions for the Voorhees metrics, given in IV as

$$F_{11} = \frac{tc(t)}{S(t)}\left(\frac{x - 2ty - S(t)}{x+1}\right)^{\delta},$$

$$F_{12} = \frac{id(t)}{S(t)}\left(\frac{x - 2ty + S(t)}{x+1}\right)^{\delta}, \tag{2.26}$$

where x, y are spherioidal coordinates,

$$\rho^2 = (x^2-1)(1-y^2), \quad z = xy, \tag{2.27}$$

and

$$cd = (1-4t^2)^{-\delta}. \tag{2.28}$$

We would like to consider in more detail some particular limits, namely, $t \to \pm\infty$ and $t \to \pm\frac{1}{2}$. The gauge $c = d$ used in IV to simplify the expression $F_{11} + itF_{12}$ is not the best choice here. One would prefer to use c, d to make the limiting values finite.

The limiting forms for $S(t)$ are

$$S(t) \sim \pm 2tr, \quad t \to \pm\infty,$$

$$S(t) = x \mp y, \quad t = \pm\tfrac{1}{2}. \tag{2.29}$$

Choosing

$$c = (1-2t)^{-\delta}, \quad d = (1+2t)^{-\delta},$$

for $t \to +\frac{1}{2}, +\infty$, and

$$c = (1+2t)^{-\delta}, \quad d = (1-2t)^{-\delta},$$

for $t \to -\frac{1}{2}, -\infty$, one obtains

$$F_{11}(\pm\infty) = \pm\tfrac{1}{2}r^{-1}(r \pm y)^{\delta}(x+1)^{-\delta},$$

$$F_{21}(\pm\infty) = -\tfrac{1}{2}ir^{-1}(r \mp xy)(r \pm y)^{\delta}(x-1)^{-\delta},$$

$$F_{11}(\pm\tfrac{1}{2}) = \pm\tfrac{1}{2}(x \mp y)^{-\delta-1}(x-1)^{\delta},$$

$$F_{21}(\pm\tfrac{1}{2}) = -\tfrac{1}{2}i(x \mp y)^{-\delta-1}(x+1)^{\delta+1}(1 \mp y). \tag{2.30}$$

3. THE NEW TRANSFORMATIONS

The vacuum Einstein equations for stationary axisymmetric spacetimes are preserved by the symmetry group **K**, consisting of the transformations $\gamma_{AB}^{(k)}$, $A,B = 1,2$, $k = 0 \pm 1, \pm 2, \cdots$. Their infinitesimal action on the potentials is given by Eq. (II.3.1):

$$\gamma_{AB}^{(k)}: N_{AB}^{(m,n)} \to N_{AB}^{(m,n)} + \gamma_{AX}^{(k)} N_{B}^{X(m+k,n)} + \gamma_{XB}^{(k)} N_{AX}^{(m,n+k)} + \gamma^{(k)XY} \sum_{s=1}^{k} N_{AX}^{(m,s)} N_{YB}^{(k-s,n)}. \tag{3.1}$$

In particular

$$N_{22}^{(k)}: N_{11}^{(m,n)} \to N_{11}^{(m,n)} + \gamma_{22}^{(k)} \sum_{s=1}^{k} N_{11}^{(m,s)} N_{11}^{(k-s,n)}. \tag{3.2}$$

Thus the knowledge of the first components $N_{11}^{(m,n)}$ alone is sufficient to describe the $\gamma_{22}^{(k)}$ transformations. In the rest of this paper we shall consider only these transformations. For convenience we simplify the notation, and from now on denote $N_{11}^{(m,n)}$ by N_{mn}. We also denote $G_{11}(s,t)$ by $G(s,t)$.

For a transformation of the form

$$\sum_{k=0}^{\infty} a_k \gamma_{22}^{(k)}, \tag{3.3}$$

where the a_k are constants, we have

$$N_{mn} \to N_{mn} + \sum_{k=0}^{\infty} a_k \sum_{s=1}^{k} N_{ms} N_{k-s,n}. \tag{3.4}$$

The exponentiation of this infinitesimal transformation into a finite one may be carried out as in III and IV. Introduce the infinite matrices

$$N = \begin{pmatrix} 0 & N_{01} & N_{02} & \cdots \\ 0 & N_{11} & N_{12} & \cdots \\ 0 & N_{21} & N_{12} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \tag{3.5}$$

$$A = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ a_1 & a_2 & a_3 & \cdots \\ a_2 & a_3 & a_4 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{3.6}$$

and the infinite-dimensional identity matrix I. Let N, $\widetilde{N}$ denote the initial and final potentials respectively. Then

$$\widetilde{N} = N + NAN + NANAN + \cdots = N + NA\widetilde{N}. \tag{3.7}$$

However, this result cannot be put to immediate use because the matrices involved are infinite-dimensional. One must look for various ways to replace A and N by finite matrices. Two examples of this have been given in III, Sec. 4, and IV, Sec. 4.

The present idea is to abandon the discrete set of basis functions N_{mn} in favor of a continuous basis $G(s,t)$. We seek transformations of the form (3.3) which are adapted to this basis, in the sense that they lead to a finite-dimensional problem.

The transformations $A^{(p)}$

Consider a transformation of the form (3.3), where the coefficients a_k are given by

$$a_k = \alpha u^k, \tag{3.8}$$

with α and u arbitrary constants. Then

$$\sum s^m t^m \widetilde{N}_{mn} = \sum s^m t^m N_{mn} + \sum s^m t^m N_{ms} \alpha u^k \widetilde{N}_{k-s,n}$$

$$\Rightarrow \widetilde{G}(s,t) = G(s,t) + \alpha G(s,u)\widetilde{G}(u,t). \tag{3.9}$$

The solution of this for $\widetilde{G}$ is

$$\widetilde{G}(s,t) = G(s,t) + \alpha G(s,u)G(u,t)/[1 - \alpha G(u,u)]. \tag{3.10}$$

More generally, we have found a series of transformations which can all be expressed in terms of $G(s,t)$ plus a few of its derivatives. We define

$$G_{ij}(s,t) = \frac{s^i t^j}{i!j!}\left(\frac{\partial}{\partial s}\right)^i \left(\frac{\partial}{\partial t}\right)^j G(s,t) = \sum_{\substack{m=i \\ n=j}}^{\infty} \frac{s^m t^n N_{mn}}{(m-i)!(n-j)!}. \tag{3.11}$$

We also define the pth rank transformation as the transformation of the form (3.3) with

$$a_k = \alpha^{(p)} \binom{k}{p} u^k, \quad k \geq p, \qquad = 0, \quad k < p, \tag{3.12}$$

where $\binom{k}{p}$ is the binomial coefficient and $\alpha^{(p)}$ and u are constants. One inserts Eq. (3.12) into Eq. (3.6), multiplies by $s^m t^n/(m-i)!(n-j)!$, and sums over m and n. This gives

$$\widetilde{G}_{ij}(s,t) = G_{ij}(s,t) + \alpha^{(p)} \sum_{k=0}^{p} G_{i,p-k}(s,u)G_{kj}(u,t) + \cdots = G_{ij}(s,t) + \alpha^{(p)} \sum_{k=0}^{p} G_{i,p-k}(s,u)\widetilde{G}_{kj}(u,t). \tag{3.13}$$

To solve Eq. (3.13) for $\widetilde{G}$, we must first set $s = u$. The result can be written

$$\sum_{k=0}^{p} M_{ik}(u)\widetilde{G}_{kj}(u,t) = G_{ij}(u,t), \qquad M_{ik}(u) \equiv \delta_{ik} - \alpha^{(p)} G_{i,p-k}(u,u). \tag{3.14}$$

Solve this system of $p+1$ linear equations by matrix inversion, and substitute the result back into Eq. (3.13) with $i = j = 0$. One gets

$$\widetilde{G}(s,t) = G(s,t) + \alpha^{(p)} \sum_{k,l=0}^{p} G_{0,p-k}(s,u) \times M^{-1}{}_{kl}(u) G_{l0}(u,t). \tag{3.15}$$

Since the Ernst potential is related to our functions by

$$\mathscr{E} = iN_{01} = i[\partial_t G(s,t)]_{s=0,t=0}, \tag{3.16}$$

we obtain

$$\widetilde{\mathscr{E}} = \mathscr{E} + i\alpha^{(p)} \sum_{k,l=0}^{p} G_{0,p-k}(0,u) M^{-1}_{kl}(u) \times [\partial_t G_{l0}(u,t)]_{t=0}, \tag{3.17}$$

which determines immediately the new solution.

Combined transformations

Successive application of these transformations with various ranks and different constants $\alpha^{(p)}$ and u may now be used to produce further new solutions. However, one should first note that the transformations $\gamma_{22}^{(k)}$ all commute, and hence the present transformations commute also. It is therefore possible to write down combined transformations in which the various $\alpha^{(p)}$'s are present simultaneously, and this can sometimes lead to shorter calculations.

For example, suppose the u's are all equal. Then the combined transformation

$$a_k = \sum_{p=0}^{N} \alpha^{(p)} \binom{k}{p} u^k \tag{3.18}$$

applied to a given solution will generate a new solution with $N+2$ additional parameters. The new solution is given by

$$\tilde{\mathscr{E}} = \mathscr{E} + i \sum_{p=0}^{N} \alpha^{(p)} \sum_{k,l=0}^{p} G_{0,p-k}(0,u) \times M^{-1}_{kl}(u)[\partial_t G_{l0}(u,t)]_{t=0}, \tag{3.19}$$

where now

$$M_{ij}(u) = \delta_{ij} - \sum_{p=i}^{N} \alpha^{(p)} G_{p-i,j}(u,u). \tag{3.20}$$

A similar procedure can be used to write combined transformations with different u parameters too. For two rank-zero transformations,

$$a_k = \alpha_1^{(0)} u_1^k + \alpha_2^{(0)} u_2^k, \tag{3.21}$$

one easily obtains that the new Ernst potential is

$$\tilde{\mathscr{E}} = \mathscr{E} + i\Delta^{-1} \sum_{j=1}^{2} \alpha_j^{(0)} A_j [\partial_t G(u_j,t)]_{t=0}, \tag{3.22}$$

where Δ is the 2×2 determinant

$$\Delta = \det[\delta_{ij} - \alpha_j^{(0)} G(u_j,u_i)] \tag{3.23}$$

and A_j is the determinant obtained by substituting the column

$$\begin{pmatrix} G(0,u_1) \\ G(0,u_2) \end{pmatrix}$$

in the determinant (3.23).

4. EXAMPLES

First we consider applications of the rank zero transformation

$$\tilde{G}(s,t) = G(s,t) + \alpha G(s,u)G(u,t)/[1-\alpha G(u,u)]. \tag{4.1}$$

From Eq. (3.16), the Ernst potential will be

$$\tilde{\mathscr{E}} = \mathscr{E} + i\alpha G(0,u)G'(u)/[1-\alpha G(u,u)]. \tag{4.2}$$

Suppose that the initial metric is a static Weyl metric. Making use of Eqs. (2.3), (2.11), and (2.18), we obtain a very general result:

$$\tilde{\mathscr{E}} = \mathscr{E} - \frac{i\alpha u^{-1} S(u)[F_{11}(u)]^2}{1-\alpha F_{11}(u)F_{21}(u)} = \left(\frac{2S^2(u) + i\alpha u[1-2uz-S(u)]\exp[2\beta(u)]}{2S^2(u) + i\alpha u[1-2uz+S(u)]\exp[2\beta(u)]} \right) \times \exp(2\chi). \tag{4.3}$$

Application to flat space

Now let us examine a few special cases. Suppose the initial metric is flat space. Then $\chi = \beta = 0$, and (dropping the tilde)

$$\mathscr{E} = \frac{2S^2 + i\alpha u(1-2uz-S)}{2S^2 + i\alpha u(1-2uz+S)}, \qquad \xi = \frac{1-\mathscr{E}}{1+\mathscr{E}} = \frac{i\alpha u S}{2S^2 + 2\alpha u(1-2uz)}. \tag{4.4}$$

In every stationary axisymmetric metric one has the coordinate freedom $z \to z + \text{const}$, and this can be used here to advantage to simplify Eq. (4.4). Let

$$z = \hat{z} + (2u)^{-1}. \tag{4.5}$$

Then, by the remark after Eq. (2.6),

$$S(u) = 2u\hat{r} \tag{4.6}$$

and we have

$$\xi = \frac{2i\alpha u^2 \hat{r}}{8u^2\hat{r}^2 - 2i\alpha u^2 \hat{z}} = \frac{\frac{1}{4}i\alpha}{\hat{r} - \frac{1}{4}i\alpha\cos\hat{\theta}}. \tag{4.7}$$

This is the well-known form for the extreme Kerr-NUT solution with $m = l = a = \frac{1}{4}\alpha$. (Extreme Kerr could, of course, be obtained by a further Ehlers transformation, $\xi \to -i\xi$.)

The rank-zero transformation applied to flat space has created an extreme Kerr-NUT particle, located at an off-center position along the axis, at $z = (2u)^{-1}$. For $u \to 0$ the transformation reduces to $\gamma_{22}^{(0)}$. Since $\gamma_{22}^{(0)}$ leaves flat space invariant, one should get in this limit the identity transformation. However, the way the solution attains this limit is rather indirect: The particle slides out along the axis to spatial infinity. Still, in a pointwise sense, the transformation does tend continuously to the identity. On the other hand, for every value of u the coordinate freedom may be used to return the particle to the origin. From this latter viewpoint, the transformation is independent of u, and hence effectively discrete.

Application to Curzon

Next, let the initial metric be the Curzon metric, i.e.,

$$\chi = -m/r, \quad \beta(t) = -mS(t)/r. \tag{4.8}$$

Now the generated solution will have two singularities present (dissimilar ones) at $z = 0$ and $z = (2u)^{-1}$. Use the coordinate change

$$z = \hat{z} + (4u)^{-1} \tag{4.9}$$

to place them symmetrically about the origin, and then transform to prolate spheroidal coordinates based on these two points:

$$x = 2ur + S(u), \quad y = 2ur - S(u), \quad \Rightarrow xy = 4uz - 1 = 4u\hat{z}. \tag{4.10}$$

The singularities are now at $x = +1, y = \pm 1$. The result from Eq. (4.3) is

$$\mathscr{E} = A\frac{(x-y)^2 + i\alpha u(1+xy-x+y)B}{(x-y)^2 + i\alpha u(1+xy+x-y)B}, \quad \xi = \frac{(x-y)^2(1-A) + i\alpha B\,[(1+xy+x-y) - A\,(1+xy-x+y)]}{(x-y)^2(1+A) + i\alpha B\,[(1+xy+x-y) + A\,(1+xy-x+y)]}, \tag{4.11}$$

where

$$A = \exp[-8mu/(x+y)], \quad B = \exp[-4mu(x-y)/(x+y)]. \tag{4.12}$$

An asymptotically flat (NUT-free) solution can be obtained from Eq. (4.11) by applying the Ehlers transformation

$$\xi \to \xi \exp(i\gamma), \quad \gamma = \tan^{-1}\left(-\frac{\alpha}{4mu}\exp(-4mu)\right). \tag{4.13}$$

Note that, in this example also, the parameter u can be entirely eliminated, by appropriately rescaling m and α. However, this will not be possible in the general case.

Application to Zipoy–Voorhees

As a final example of application of the rank-zero transformation, let the initial metric be the Zipoy–Voorhees metric. From Eqs. (2.26), (4.3) we find

$$\tilde{\mathscr{E}} = \left(\frac{x-1}{x+1}\right)^{\delta}\left(1 - \frac{2i\alpha c^2(u)uS(u)(x-2uy-S(u))^{2\delta}}{2S^2(u)(x^2-1)^{\delta} + i\alpha u c^2(u)(1-2uxy+S(u))(x-2uy-S(u))^{2\delta}}\right). \tag{4.14}$$

Here the arbitrary gauge function $c(u)$ has been retained. However, for any fixed value of u, the effect of $c(u)$ may be absorbed into a redefinition of α. Consequently, the gauge freedom does not lead to further solutions. Instead, it may be used as in Sec. 2 to simplify various limits.

We expect to find the original rod singularity lying between $z = \pm 1$, and the point singularity at $z = (2u)^{-1}$. Interesting cases are therefore $u = \pm \frac{1}{2}$, when the point coincides with one end of the rod, and $u = \pm \infty$ when it moves to the origin. Using the limits found in Eqs. (2.29), (2.30), we obtain for $u = +\frac{1}{2}$,

$$\mathscr{E} = \left(\frac{x-1}{x+1}\right)^{\delta}\left[\frac{(x-y)^{2\delta+2} - \frac{1}{4}i\alpha(x^2-1)^{\delta}(x-1)(1+y)}{(x-y)^{2\delta+2} + \frac{1}{4}i\alpha(x^2-1)^{\delta}(x+1)(1-y)}\right]. \tag{4.15}$$

For $u = +\infty$, we find

$$\mathscr{E} = \left(\frac{x-1}{x+1}\right)^{\delta}\left[\frac{r^2(r-y)^{\delta} - \frac{1}{4}i\alpha(r+y)^{\delta}(r+xy)}{r^2(r-y)^{\delta} + \frac{1}{4}i\alpha(r+y)^{\delta}(r-xy)}\right], \quad \xi = \frac{Br^2(r-y)^{\delta} + \frac{1}{4}i\alpha(Ar-Bxy)(r+y)^{\delta}}{Ar^2(r-y)^{\delta} + \frac{1}{4}i\alpha(Br-Axy)(r+y)^{\delta}}, \tag{4.16}$$

where

$$A = \tfrac{1}{2}[(x+1)^{\delta} + (x-1)^{\delta}], \quad B = \tfrac{1}{2}[(x+1)^{\delta} - (x-1)^{\delta}]. \tag{4.17}$$

The case $u = -\frac{1}{2}, -\infty$ are obtained by replacing $y \to -y$ and $\alpha \to -\alpha$ in the above. Physically this amounts to turning the source upside down along with its spin. The limits $u = \pm\infty$, when the particle approaches the origin from above or below, are not identical as one might have naively supposed, indicating that the source somehow remains asymmetrical.

Combined transformations

Next we consider examples of applications of the combined transformations.

Application to flat space of the combined rank-zero and rank-one transformations gives a two-parameter asymptotically flat solution which generalizes the extreme Kerr solution. Since the solution has been given in Ref. 7, we do not repeat it here. Instead, we devote the rest of the section to the study of applications of two zero rank transformations with different u parameters. Acting on flat space the transformation creates the spacetime due to two Kerr-NUT particles located on the axis at $z = (2u_1)^{-1}$ and $z = (2u_2)^{-1}$. By changing z,

$$z \to z + (u_1 + u_2)/4u_1u_2, \tag{4.18}$$

and defining prolate spheroidal coordinates by

$$S(u_1) = [(u_1 - u_2)/2u_2](x+y), \quad S(u_2) = [(u_1 - u_2)/2u_1](x-y), \tag{4.19}$$

one obtains the solution with Ernst ξ-potential

$$\xi = \frac{i(x^2-y^2)[\alpha_1(x-y) + \alpha_2(x+y)] + 2\alpha_1\alpha_2 y(y^2-1)}{(x^2-y^2)^2 - i[\alpha_1(x-y)^2(xy+1) + \alpha_2(x+y)^2(xy-1)] + \alpha_1\alpha_2(1-y^4)}. \tag{4.20}$$

This is a special case of the generalized Tomimatsu–Sato metric discovered earlier, Eq. (IV. 4.30), in which $p = 0, q = 1$, $\gamma = -\pi$, and

$$\alpha = \tfrac{1}{2}(1/\alpha_1 - 1/\alpha_2), \quad \beta = -\tfrac{1}{2}(1/\alpha_1 + 1/\alpha_2). \tag{4.21}$$

The final example is the application of the transformation Eq. (3.22) to the Zipoy–Voorhees metrics. Since the resulting experessions are rather lengthy, we give them only in the limit $u_1 = \frac{1}{2}, u_2 = -\frac{1}{2}$, where they simplify considerably. Using Eqs. (2.30), we have

$$\mathscr{E} = \left(\frac{x-1}{x+1}\right)^{\delta} \times \frac{2i(x-1)^{2\delta}(x^2-y^2)[\alpha_2(x-y)^{2\delta+1} - \alpha_1(x+y)^{2\delta+1}] + 4\alpha_1\alpha_2 x(x-1)^{2\delta}(x^2-1)^{\delta+1}}{(x^2-y^2)^{2\delta+2} - i(x^2-1)^{\delta}(x+1)[\alpha_1(y-1)(x+y)^{2\delta+2} + \alpha_2(y+1)(x-y)^{2\delta+2}] - \alpha_1\alpha_2(x^2-1)^{2\delta+1}(x+1)^2}, \tag{4.22}$$

5. MULTIPOLE MOMENTS

Consider a stationary asymptotically flat vacuum solution. Let r be a coordinate which is asymptotically the Euclidean radius. Then the Ernst potential with respect to the stationary Killing field is of the form

$$\mathscr{E} = 1 + Ar^{-1} + Br^{-2} + O(r^{-3}), \tag{5.1}$$

where A is a real constant, B and the coefficients of higher order terms will be appropriate smooth complex functions of polar angle. Conversely, a stationary solution whose Ernst potential satisfies these conditions will be asymptotically flat. The transformations introduced in Sec. 3 may easily be shown to preserve all of these conditions, except for the reality of A. Since this can be cured by a subsequent Ehlers transformation, we are always led to a new asymptotically flat solution.

Now let us examine the far-field limit of the transformations in more detail, to determine their effect on the multipole moments. Consider $G(s,t)$ for the general Weyl solution. As $r \to \infty$, $\chi \to 0$. From Eq. (2.19), $\beta(t) \to 0$ also. Then, from Eq. (2.18),

$$G(s,t) = -\tfrac{1}{4}ir^{-1}(1-\cos\theta) + O(r^{-2}). \tag{5.2}$$

Note that the flat-space term dominates over any nonflat ones. Hence, as a preliminary study of the effect of the transformations on the multipole moments, we examine their effect on flat space.

In Eq. (3.13), the dominant effect will come from the term with the least number of G's, namely the term linear in $\alpha^{(p)}$:

$$\tilde{G}(s,t) \approx G(s,t) + \alpha^{(p)} \sum_{k=0}^{p} G_{0,p-k}(s,u)G_{k0}(u,t). \tag{5.3}$$

Using Eq. (3.11) and Leibnitz's rule, we have

$$\tilde{G}(s,t) \approx G(s,t) + \frac{\alpha^{(p)}u^p}{p!}\left(\frac{\partial}{\partial u}\right)^p [G(s,u)G(u,t)]. \tag{5.4}$$

We now insert the flat-space value for $G(s,t)$ from Eq. (2.18) and use Eq. (3.16) to get eht Ernst potential:

$$\tilde{\mathscr{E}} = 1 - \frac{i\alpha^{(p)}u^p}{p!}\left(\frac{\partial}{\partial u}\right)^p\left(\frac{u}{S(u)}\right). \tag{5.5}$$

Expand for large r,

$$S^{-1}(u) = \sum_{n=0}^{\infty} (2ur)^{-n-1}P_n(\cos\theta), \tag{5.6}$$

$$uS^{-1}(u) = (2r)^{-1} + (4r^2u)^{-1}P_1(\cos\theta) + \cdots. \tag{5.7}$$

For $p = 0$, the leading term is the r^{-1} monopole. That is, the transformation produces Schwarzschild mass. For $p \geqslant 1$, the leading nonzero term will always be the dipole, and each transformation in the series produces angular momentum, but does not change the mass.

Generation of pure multipoles

It would be even more desirable to have a somewhat different sequence of transformations, say $D^{(p)}$, which would leave the first p multipoles unchanged, and change the rest starting with the $(p+1)$th. Using such a sequence, one could adjust each multipole moment in turn, and thus routinely produce exact stationary solutions with any prescribed set of moments.

Such a sequence of transformations is not hard to find. They will be simple linear combinations of the $A^{(p)}$, chosen to produce a cancellation in the generated dipole, quadrupole, etc. Proceeding by induction, one can show that the transformations with

$$a_k = \delta^{(p)}\binom{k+p-1}{p}u^k \tag{5.8}$$

are the ones we seek.

In fact, the new transformations $D^{(p)}$ would have arisen naturally to begin with if we had used inverse parameters for the generating functions, e.g.,

$$\bar{G}(s,t) = G(s^{-1},t^{-1}). \tag{5.9}$$

To illustrate this point, let

$$v = (4u)^{-1}, \quad R = r^{-1}.$$

Then

$$S(u,r) = (2vR)^{-1}S(v,R), \tag{5.10}$$

$$uS^{-1}(u,r) = \tfrac{1}{2}RS^{-1}(v,R) = \tfrac{1}{2}R\sum(2vR)^nP_n(\cos\theta), \tag{5.11}$$

Instead of Eq. (5.4), we would have been led to the transformation

$$\tilde{\mathscr{E}} = 1 - \frac{i\delta^{(p)}v^p}{p!}\left(\frac{\partial}{\partial v}\right)^p\left(\frac{R}{2S(v,R)}\right)$$

$$= 1 - \tfrac{1}{4}i\delta^{(p)}V^p(2R)^{p+1}P_p(\cos\theta) + O(R^{p+2}). \tag{5.12}$$

which has a 2^p-multipole as its leading term.

Note added in proof: Due to a change in method of composition, we have to use a new notation for the potentials $N_{AB}^{(m,n)}$ and transformations $\gamma_{AB}^{(k)}$. Formerly m, n, k were placed above the main character.

[1]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977) (I).
[2]W. Kinnersley and D.M. Chitre, J. Math. Phys. **18**, 1538 (1977) (II).
[3]W. Kinnersley and D.M. Chitre, J. Math. Phys. **19**, 1926 (1978) (III).
[4]W. Kinnersley and D.M. Chitre, J. Math. Phys. **19**, 2037 (1978) (IV).
[5]C. Hoenselaers, J. Math. Phys. **20**, xxxx (1979) (V).
[6]W. Kinnersley and D.M. Chitre, Phys. Rev. Lett. **40**, 1608 (1978).
[7]C. Hoenselaers, W. Kinnersley, and B.C. Xanthopoulos, Phys. Rev. Lett. **42**, 481 (1979).

Charged fluid sphere in general relativity

D. N. Pant and A. Sah

Department of Mathematics, Kumaun University, Naini Tal 263002, India
(Received 9 October 1978)

An analytic solution of the relativistic field equations is obtained for a static, spherically symmetric distribution of charged fluid. The arbitrary constants are determined by matching it with the Reissner–Nordström solution over the boundary. The distribution behaves like a charged perfect gas. As a particular case a solution for a spherical distribution of charged incoherent matter is deduced where the charge density and the mass density are equal in magnitude. In the absence of the charge, the solution reduces to Tolman's solution VI with $B = 0$.

1. INTRODUCTION

On account of the occurrence of two singularities in the Reissner–Nordström metric corresponding to two nonzero finite values of r, it is generally suggested that the collapse of a spherical material distribution to a point singularity can be avoided if the matter is accompanied by some charge. In such a model the gravitational attraction is balanced by the electrical repulsion together with the hydrostatic pressure of the matter. Therefore, to start with, it is desirable to construct new solutions of the field equations in general relativity representing the internal field of charged distributions. The external field of such distributions (spherically symmetric) will be described by the Reissner–Nordström solution. In this paper we have solved the coupled Einstein–Maxwell field equations of general relativity for a static, spherically symmetric distribution of charged fluid. The fluid model has nonnegative expressions for the mass density and pressure. The magnitude of these parameters and the charge density is maximum at the center.

The combined Einstein–Maxwell field equations are (using the geometric units $G = c = 1$)

$$R_i^j - \tfrac{1}{2}Rg_i^j = -8\pi T_i^j, \tag{1}$$

where the energy–momentum tensor T_i^j is given by

$$T_i^j = M_i^j + E_i^j, \tag{2a}$$

$$M_i^j = (\rho + p)u_i u^j - pg_i^j, \tag{2b}$$

$$E_i^j = \frac{1}{4\pi}(-F_{i\alpha}F^{j\alpha} + \tfrac{1}{4}g_i^j F_{\alpha\beta}F^{\alpha\beta}). \tag{2c}$$

u^i is the timelike four-velocity unit vector and F_{ij} satisfy Maxwell's equations

$$F_{ij,k} + F_{jk,i} + F_{ki,j} = 0, \tag{3a}$$

$$\frac{\partial}{\partial x^j}(\sqrt{-g}F^{ij}) = 4\pi\sqrt{-g}J^i, \tag{3b}$$

J^i being the charge-current vector. We consider the fluid to be of null conductivity, so that, if σ denotes the charge density,

$$J^i = \sigma u^i. \tag{3c}$$

The static, spherically symmetric space–time considered in our investigation is described by the line element

$$ds^2 = -e^{\lambda(r)}dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2) + e^{\nu(r)}dt^2. \tag{4}$$

Since the field is static, we have

$$u^i = (0,0,0,e^{-\nu/2}) \tag{5}$$

and

$$J^1 = J^2 = J^3 = 0. \tag{6}$$

Using (6) and the condition for spherical symmetry we find from (3a), (3b) that the only nonvanishing components of F_{ij} are F_{14} and F_{41}. Hence, in view of (2a), (2b), (2c), (4), and (5), the set of field equations (1) reduces to the following:

$$\begin{aligned} &e^{-\lambda}\left(\frac{\nu'}{r} + \frac{1}{r^2}\right) - \frac{1}{r^2} = 8\pi(p - k), \\ &e^{-\lambda}\left(\frac{\nu''}{2} + \frac{\nu'^2}{4} + \frac{\nu' - \lambda'}{2r} - \frac{\lambda'\nu'}{4}\right) = 8\pi(p + k), \\ &e^{-\lambda}\left(\frac{\lambda'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2} = 8\pi(\rho + k), \end{aligned} \tag{7}$$

where the prime denotes differentiation with respect to r and

$$k = -\frac{1}{8\pi}F_{41}F^{41}. \tag{8}$$

2. A SOLUTION OF THE EQUATIONS

An analytic solution of these equations is obtained. The arbitrary constants have been determined by matching it with the Reissner–Nordström solution

$$e^{-\lambda} = e^{\nu} = 1 - \frac{2m}{r} + \frac{\epsilon^2}{r^2}, \quad F_{14} = \frac{\epsilon}{r^2} \tag{9}$$

over the boundary. The solution is

$$\begin{aligned} &e^{-\lambda} = c, \quad e^{\nu} = Ar^{2n}, \\ &\rho = \frac{1}{16\pi r^2}\{1 - c(n-1)^2\}, \\ &p = \frac{1}{16\pi r^2}\{c(n+1)^2 - 1\}, \\ &k = \frac{1}{16\pi r^2}\{1 + c(n^2 - 2n - 1)\}, \end{aligned} \tag{10a}$$

 0022-2488/79/122537-03$01.00

where

$$A = r_0^{-2n}\left(1 - \frac{2m_0}{r_0} + \frac{\epsilon_0^2}{r_0^2}\right), \tag{10b}$$

$$c = \left(1 - \frac{2m_0}{r_0} + \frac{\epsilon_0^2}{r_0^2}\right) = \left(1 - \frac{2\epsilon_0^2}{r_0^2}\right)(1 + 2n - n^2)^{-1}. \tag{10c}$$

Here ϵ_0 is the electric charge and m_0 the mass, as measured by an external observer, of a fluid sphere of radius r_0. From (3b), (3c), (8), and (10a) we obtain for the charge density

$$\sigma = \pm \frac{1}{4\pi r^2}\left(\frac{c}{2}\left[1 + c\sqrt{n^2 - 2n - 1}\right]\right)^{1/2}. \tag{11}$$

In order that ρ and p be positive throughout we have the following conditions on c:

$$c \leqslant (n-1)^{-2}, \tag{12a}$$

$$c \geqslant (n+1)^{-2}. \tag{12b}$$

σ will always be real for $1 - (2)^{1/2} \geqslant n \geqslant 1 + (2)^{1/2}$. For $1 - (2)^{1/2} < n < 1 + (2)^{1/2}$, σ will be real under the condition

$$c \leqslant (-n^2 + 2n + 1)^{-1}. \tag{12c}$$

The principle of causality $dp/d\rho \leqslant 1$ further restricts c by

$$c \leqslant (n^2 + 1)^{-1}. \tag{12d}$$

The perfect fluid property $\rho \geqslant 3p$ puts a stronger condition than (12d) viz.,

$$c \leqslant (n^2 + n + 1)^{-1}. \tag{12d'}$$

Since c cannot exceed unity, (12b) restricts n to $n \leqslant -2$ or $n \geqslant 0$. But (12a) and (12b) together will hold only when $n \geqslant 0$. Also from (10c), c becomes infinite for $n = 1 + (2)^{1/2}$ and negative for $n > 1 + (2)^{1/2}$. We conclude that (10) will represent a realistic model of the charged fluid sphere for $0 \leqslant n < 1 + (2)^{1/2}$ if

$$(n+1)^{-2} \leqslant c \leqslant (-n^2 + 2n + 1)^{-1}, \quad \text{for} \quad 0 < n \leqslant \tfrac{1}{2}, \tag{13}$$

$$(n+1)^{-2} \leqslant c \leqslant (n^2 + n + 1)^{-1}, \quad \text{for} \quad \tfrac{1}{2} \leqslant n < 1 + (2)^{1/2}.$$

The electrostatic field inside the sphere is described by

$$F_{14} = \pm\left(\frac{A}{2c}(1 + c\sqrt{n^2 - 2n - 1})\right)^{1/2} r^{n-1}. \tag{14}$$

The mass density, pressure, and the magnitude of the charge density decrease as the inverse square of the distance from the center and become infinite at the center. Also, from (10a) the equation of state is

$$p \propto \rho, \tag{15}$$

which implies that our distribution behaves like a perfect gas. Further, it is to be noted that the solution provides an exact relationship between the mass m_0 and the charge ϵ_0 of the distribution through (10c).

The surviving components of the Riemann–Christoffel curvature tensor R_{hijk} for the metric (4) in view of (10a), are given by

$$R_{1414} = nA(1-n)r^{2(n-1)},$$

$$R_{2323} = r^2(c + 1\sin^2\theta - 1),$$

$$R_{2424} = -ncAr^{2n}, \tag{16}$$

$$R_{3434} = -ncAr^{2n}\sin^2\theta.$$

Let $\lambda^{h}_{(a)}$ denote an orthonormal tetrad associated with an event in space–time. We choose

$$\lambda^{h}_{(a)} = \text{diagonal}\left(-\sqrt{-g^{11}}, -\sqrt{-g^{22}}, -\sqrt{-g^{33}}, \sqrt{g^{44}}\right). \tag{17}$$

The physical components $R_{(abcd)}$ of the curvature tensor are defined by

$$R_{(abcd)} = \lambda^{h}_{(a)}\lambda^{i}_{(b)}\lambda^{j}_{(c)}\lambda^{k}_{(d)}R_{hijk}. \tag{18}$$

The nonvanishing components are

$$R_{(1414)} = \frac{n(1-n)c}{r^2},$$

$$R_{(2323)} = \frac{(c + 1 - \text{cosec}^2\theta)}{r^2}, \tag{19}$$

$$R_{(2424)} = R_{(3434)} = \frac{-nc}{r^2}.$$

Since $R_{(abcd)} \to 0$ as $r \to \infty$, the space–time is asymptotically flat.

3. PARTICULAR CASES

A. Charged disordered radiation

The equation of state for disordered radiation is $\rho = 3p$. Hence for this type of distribution our solution (10a) becomes,

$$e^{-\lambda} = (n^2 + n + 1)^{-1}, \quad e^{\nu} = Ar^{2n},$$

$$\rho = 3p = \frac{3}{16\pi r^2}\frac{n}{n^2 + n + 1}, \tag{20}$$

$$\sigma = \pm \frac{1}{4\sqrt{2}\pi r^2}\frac{\sqrt{n(2n-1)}}{(n^2 + n + 1)}.$$

This solution represents a real distribution of charged disordered radiation for $n \geqslant \frac{1}{2}$. Another interesting case is the extreme case of very relativistic gas where the speed of sound attains the maximum value viz., the speed of light. This is characterized by $dp/d\rho = 1$, which leads to $c = (n^2 + 1)^{-1}$. Thus, our solution becomes

$$e^{-\lambda} = (n^2 + 1), \quad e^{\nu} = Ar^{2n},$$

$$\rho = p = \frac{n}{8\pi r^2(n^2 + 1)}, \tag{21}$$

$$\sigma = \frac{\sqrt{n(n-1)}}{4\pi r^2(n^2 + 1)},$$

which is valid for $n \geqslant 1$.

B. Charged dust case

Here we consider a distribution of charged incoherent matter where the attracting gravitational field of force is balanced by the repulsive electrostatic field of force. In such a case $p = 0$ and the solution reduces to

$$e^{-\lambda} = (n+1)^{-2}, \quad e^{\nu} = Ar^{2n},$$

$$\rho = \frac{n}{4\pi r^2(n+1)^2}, \quad \sigma = \pm \frac{n}{4\pi r^2(n+1)^2}, \tag{22}$$

where A is given by (10b). We find that the mass density and the charge density are equal in magnitude. If $N(r)$ denotes the particle number density of the distribution and $\epsilon(r)$, $m(r)$ denotes the charge and mass associated with a particle, we have the relation

$$\rho = m(r)N(r), \quad \sigma = \epsilon(r)N(r). \tag{23}$$

We arrive at the conclusion

$$\frac{\epsilon^2(r)}{m^2(r)} = 1. \tag{24}$$

Thus, the solution (22) is in agreement with the Newtonian theory, which permits the stability of a static distribution of charged incoherent mass particles only when (24) holds. The solution has a physical character for all positive values of the parameter n. Apart from this, in view of (10c), we have the relations

$$\epsilon_0 = \pm \frac{n}{n+1} r_0, \quad m_0 = \frac{n}{n+1} r_0, \tag{25}$$

giving exact expressions for the total charge and mass of the sphere in terms of its radius.

C. Uncharged fluid sphere

From (11) $\sigma = 0$ if $c = (1 + 2n - n^2)^{-1}$ which in view of (10c) makes $\epsilon = 0$. Our solution then takes the form:

$$e^{\lambda} = 2 - (n-1)^2 = \left(1 - \frac{2m_0}{r_0}\right)^{-1},$$

$$e^{\nu} = Ar^{2n} = r_0^{2n}\left(1 - \frac{2m_0}{r_0}\right)r^{2n}, \tag{26}$$

$$\rho = \frac{1}{8\pi r^2}\frac{n(2-n)}{2-(n-1)^2}, \quad p = \frac{1}{8\pi r^2}\frac{n^2}{2-(n-1)^2},$$

which is Tolman's solution VI with $B = 0$ (Ref. 1). Here the condition $p \geqslant 0$, $\rho \geqslant 0$ together with the principle of causality demand $0 \leqslant n \leqslant 1$, which implies that the fluid model given by (26) has a maximum mass (as measured by an external observer) of one fourth the fluid radius. Moreover, if we impose the stronger condition $\rho \geqslant 3p$, the solution is restricted to $0 \leqslant n \leqslant \frac{1}{2}$, thereby reducing the mass range to $0 \leqslant m_0/r_0 \leqslant \frac{3}{14}$. This solution, like that of Adler, [2] has no surface of zero pressure.

ACKNOWLEDGMENT

The authors express their gratitude to the referee for his valuable suggestions in the preparation of the paper. Their thanks are also due to Dr. S.D. Sinvhal for providing necessary literature at U.P. State Observatory, Naini Tal.

[1] R.C. Tolman, Phys. Rev. **55**, 364 (1939).
[2] R.J. Adler, J. Math. Phys. **15**, 727 (1974).

Even parity junction conditions for perturbations on most general spherically symmetric space–times

U.H. Gerlach [a)]
Department of Mathematics, Ohio State University, Columbus, Ohio 43210

U.K. Sengupta [b)]
Department of Physics, Ohio State University, Columbus, Ohio 43210

(Received 30 January 1979; accepted for publication 25 June 1979)

A new highly efficient and versatile general relativistic perturbational formalism for general matter occupied spherically symmetric space–times is developed. The perturbations are geometrical objects on the two dimensional totally geodesic submanifold spanned by the radial and time coordinates. The geometrical objects are "gauge invariant" scalars, vectors, and tensors which are independent of infinitesimal coordinate transformations on the background space–time. This article gives the even parity gauge invariant perturbation objects for arbitrary background scalars, vectors, and symmetric tensors on a spherically symmetric space–time. In particular, metric, matter, first and second fundamental forms, as well as vacuum-matter interface gauge invariant perturbations for a collapsing star are given. In addition four even parity continuity conditions across discontinuous timelike hypersurfaces are given. Two are conditions on the metric gauge invariants, one is a condition on the perturbation away from the spherical contour of the interface, and the fourth couples that contour perturbation to the metric gauge invariants.

I. INTRODUCTION AND SUMMARY

Junction conditions in general relativity arise from the answer to the following question: Given a space–time having a metric and stress energy as dictated by the Einstein field equation and given a three-dimensional hypersurface, how is the metric (and the stress–energy) tensor to be continued across the hypersurface? The answer is well known; indeed, if the stress energy has only a finite jump discontinuity in a well-defined sense, then the intrinsic metric (first fundamental form) of the surface of discontinuity, its extrinsic curvature (second fundamental form) as well as the flow of energy–momentum across this surface are each continuous across the hypersurface.[1] If one asks now the same question about a space–time perturbed to first order away from, say, a general spherically symmetric background the answer is, of course, the same. The junction conditions for the perturbations themselves can be roughly obtained in two ways:

(a) Perturb the background junction conditions and obtain thereby the junction conditions for the perturbations.

(b) Consider the Einstein field equations linearized around, say, some spherically symmetric space time. If certain of its coefficients are allowed to have discontinuities to reflect those of the background, determine the necessary junction conditions that have to be obeyed by the solution to the set of linear partial differential equations.

This paper uses the first of these two ways to arrive at the junction conditions.

The least reason for considering even perturbations in general and their junction conditions in particular is by no means the fact that asymmetrically pulsating stars undergoing gravitational collapse. They are prime candidates for the source of gravitational radiation which present day and future detectors are hoped to receive.[2] One of the types of gravitational collapse that is of particular astrophysical interest is a slightly aspherical general relativistic version of the type first considered by Colgate and White[3] or May and White.[4] Within this context it is our purpose to make precise general relativistic perturbational statements about such astrophysical events. For odd parity perturbations it is straightforward to write down a master scalar equation from whose solution all odd parity metric perturbation are derived.[5] These perturbations are, however, only associated with viscous or shear stresses and not with any pressure (even parity) perturbations.

Viscous stresses promise to play a particularly interesting role in slightly aspherical gravitational collapse.[6] Indeed indications are that with the existence of viscosity due to coherent neutrino scattering, gravitational radiation has a significant competitor for damping out asymmetries during gravitational collapse.[7]

Asymmetric (even parity) pressure perturbations are, however, also expected to play a significant role in the production of gravitational radiation. On a generic spherically symmetric space–time even parity perturbation theory has as yet not been reduced to a master scalar equation for the gravitational degree of freedom in a way that odd parity perturbation theory has. Nevertheless, a set of coupled equations for the even parity gauge invariant[8] perturbation object does exist[5] and their relevant junction conditions at a discontinuous (e.g., matter-vacuum) interface can, as we shall presently see, be written down.

The results of this article can be stated as follows: Consider a spherically symmetric space time with metric

$$g_{\mu\nu}dx^{\mu}\,dx^{\nu} = g_{AB}dx^{A}dx^{B} + r^{2}(x^{C})[d\theta^{2} + \sin^{2}\theta\, d\varphi^{2}] \tag{1.1}$$

and stress–energy tensor

[a)] On leave of absence at Center for Theoretical Physics, University of Texas, Austin, TX 78712

[b)] Now at Honeywell Electro Optics Center, Lexington, Mass. 02173

 0022-2488/79/122540-07$01.00

$$t_{\mu\nu}\,dx^{\mu}\,dx^{\nu} = t_{AB}dx^{A}\,dx^{B} + \tfrac{1}{2}t_{a}{}^{a}r^{2}[d\theta^{2} + \sin^{2}\theta\,d\varphi^{2}] \tag{1.2}$$

Consider an even parity perturbation of the background fields, Eqs. (1.1), (1.2), and of the vector field

$$n_{\mu}\,dx^{\mu} = (f_{,\alpha}\,f_{,\beta}g^{\alpha\beta})^{-1/2}f_{,\mu}dx^{\mu} = n_{A}\,dx^{A},$$

normal to the concentric level surfaces (three-dimensional submanifolds) of the scalar function $f(x^{\mu}) = f(x^{C})$, $C = 0,1$. These perturbations are expressed in terms of the spherical harmonic $Y(\theta,\varphi)$, its derivatives $Y_{,a}$ and $Y_{,a:b}$ (suppress angular integers l and m) and the metric γ_{ab} on the unit two sphere as follows:

metric:

$$\begin{aligned} &h_{\mu\nu}dx^{\mu}\,dx^{\nu} \\ &= h_{AB}(x^{C})Ydx^{A}\,dx^{B} + h_{A}(x^{C})Y_{,a}(dx^{A}\,dx^{a} + dx^{a}dx^{A}) \\ &\quad + [r^{2}K(x^{C})Y\gamma_{ab} + r^{2}G(x^{C})Y_{,a:b}]dx^{a}dx^{b}; \end{aligned} \tag{1.3}$$

matter[9]:

$$\begin{aligned} \Delta t_{\mu\nu}dx^{\mu}\,dx^{\nu} &= \Delta t_{AB}\,Ydx^{A}dx^{B} + \Delta t_{A}\,Y_{,a}(dx^{A}dx^{a} + dx^{a}dx^{A}) \\ &\quad + [r^{2}\Delta t^{1}Y\gamma_{ab} + \Delta t^{2}Y_{,a:b}]dx^{a}\,dx^{b}; \end{aligned}$$

surface:

$$\begin{aligned} &\Delta n_{\mu}dx^{\mu} = \Delta n_{A}\,Ydx^{A} + \Delta nY_{,a}dx^{a}, \\ &\Delta f'(x^{\mu}) = \Delta f(x^{C})Y(\theta,\varphi). \end{aligned}$$

Construct the corresponding gauge invariant metric, matter and surface perturbations

$$k_{AB} = h_{AB} - p_{A|B} - p_{B|A}, \tag{1.4}$$

$$k = K - 2v^{A}p_{A},$$

where

$$v_{A} = r_{,A}/r \tag{1.5}$$

and

$$\left.\begin{aligned} T_{AB} &= \Delta t_{AB} - t_{AB}{}^{|C}p_{C} - t_{A}{}^{C}p_{C|B} - t_{B}{}^{C}p_{C|A} \\ T_{A} &= \Delta t_{A} - t_{A}{}^{C}p_{C} - r^{2}(t_{a}{}^{a}/4)G_{,A} \\ T^{1} &= \Delta t^{2} - (p^{C}/r^{2})(r^{2}t_{a}{}^{a}/2)_{,C} \\ T^{2} &= \Delta t^{2} - (r^{2}t_{a}{}^{a}/2)G \end{aligned}\right\} \text{(matter)},$$

$$\left.\begin{aligned} N_{A} &= \Delta n_{A} - n_{A|C}\,p^{C} - n^{C}p_{C|A} \\ N &= \Delta n - n_{C}\,p^{C} \\ F &= \Delta f - f_{,C}\,p^{C} \end{aligned}\right\} \text{(surface)}, \tag{1.6}$$

where

$$p_{C} = h_{C} - \tfrac{1}{2}r^{2}G_{,C}. \tag{1.7}$$

As in the odd parity case,[5] note that once these gauge invariants are given, all perturbed quantities $h_{AB},\ldots,\Delta t_{AB},\ldots,\Delta f$ can be obtained by subjecting p_{C} and G to (a total of three) gauge conditions. The well-known Regge–Wheeler gauge, for example consists of $p_{C} = G = 0$.

For $l = 0,1,2,\ldots$ the relevant equations for these (even parity) gauge invariants have been given elsewhere[5] and are not presented here. Their junction conditions consist of the following quantities being continuous across a spherically symmetric hypersurface across which the stress–energy tensor is discontinuous:

$$n_{C}v^{C}\,k_{AB}i^{AB} + n_{C}{}^{|C}k; \tag{1.8a}$$

$$n^{C}k_{CB}i^{\ B}_{A}; \tag{1.8b}$$

$$N; \tag{1.8c}$$

$$\begin{aligned} &(D_{A}a^{A} - a_{A}a^{A} - 2v_{A}a^{A})v_{C}n^{C}N + \\ &+ k_{BD}n^{B}n^{D}(v_{A}n^{A} + \tfrac{1}{2}n_{A}{}^{|A})v_{C}n^{C} + \\ &+ [k_{,A}n^{A} - \tfrac{1}{2}n^{E}(k_{EA|B} + k_{EB|A} - k_{AB|E})]v_{C}n^{C} \\ &- [n_{C}{}^{|C}n_{B}{}^{|B} + (v_{A}n^{A} + \tfrac{1}{2}n_{A}{}^{|A})_{|B}n^{B}]k. \end{aligned} \tag{1.8d}$$

The deformation of the surface Σ is given by the gauge invariant vector

$$N_{A} = (N_{,B} + a_{B}N)i^{B}_{A} + \tfrac{1}{2}n^{B}n^{D}k_{BD}n_{A}. \tag{1.9}$$

Here $a_{B} \equiv n_{B|C}n^{C}$ is the acceleration of an observer on Σ, $i_{AB} \equiv g_{AB} - n_{A}n_{B}$ is the metric intrinsic to Σ, and

$$D_{A}a^{A} = a^{A}{}_{|B}i_{A}{}^{B}$$

is the divergence of a^{A} on Σ.

Even parity metric perturbations and their junction conditions have not yet been formulated in terms of a single master scalar wave function the way odd parity perturbations have. Nevertheless, comparing the derivations of even and odd parity junction conditions one can say this:

Eq. (1.8a), obtained from perturbations in the first fundamental form of Σ and Eq. (1.8b), obtained from perturbations ($\propto Y\gamma_{ab}$) of the second fundamental form, constitute most probably the intrinsic and the normal components of the gradient of the yet-to-be-found even parity master scalar wave function. Equations (1.8c) and (1.8d) are continuity conditions imposed on the perturbation of the star's surface. It is evident that (1.8d) expresses the fact that these surface perturbations are coupled to the perturbations in the gravitational field itself.

Equation (1.9) expresses the surface perturbation vector N_{A} in terms of the surface and metric perturbations N and k_{AB}.

II. REDUCTION OF TENSOR FIELDS ON A SPHERICALLY SYMMETRIC SPACE TIME

A general spherically symmetric background space time M has a metric of the form

$$\begin{aligned} &g_{\mu\nu}dx^{\mu}\,dx^{\nu} \\ &\quad = g_{AB}(x^{C})dx^{A}dx^{B} + r^{2}(x^{C})[d\theta^{2} + \sin^{2}\theta\,d\varphi^{2}]. \end{aligned} \tag{2.1}$$

The functions $r(x^{C})$ and $g_{AB}(x^{C})$ are scalar and tensor fields on the totally geodesic two-dimensional space–time M^{2} spanned by the two coordinates x^{C} ($C = 0,1$). The vector field

$$v_{A} = r_{,A}/r \tag{2.2}$$

is also on this submanifold.

Let

$$f(x^{\mu}) = f(x^{C}) \tag{2.3}$$

be a scalar function independent of the angular variables. Its contours are spherically symmetric 3-manifolds in M or simply one dimensional manifolds in M^{2}.

The first and second fundamental forms of these 3-manifolds have the form

$$i_{AB} = g_{AB} - n_{A}n_{B}, \tag{2.4a}$$

$$i_{Aa} = 0,$$

$$i_{ab} = g_{ab} = r^{2}\gamma_{ab} \tag{2.4b}$$

and

$$e_{AB} = n_{A|B} - n_{A|C}n^C n_B = n_C{}^{|C} i_{AB} \,, \tag{2.5a}$$

$$e_{Aa} = 0 \,,$$

$$e_{ab} = \tfrac{1}{2} e_c{}^c g_{ab} = v_C n^C g_{ab} \,. \tag{2.5b}$$

Latin lower cases a,b,c refer to the angular coordinates. The symbol γ_{ab} is the metric on the unit two sphere. The vertical slash | in Eq. (2.5a) and in all subsequent equations, refers to the covariant derivative with respect to g_{AB} on M^2. The second equality in Eq. (2.5a),

$$e_{AB} = e_C{}^C i_{AB} \,,$$

is true for any symmetric tensor with the property $e_{AB}n^B = 0$ on M^2. The partial trace $e_c{}^c = e_2{}^2 + e_3{}^3 = 2v_C n^C$ in Eq. (2.5b) is a scalar on M^2.

It is clear that $f(x^C), i_{AB}, e_{AB}, n^C_{|C}, v_C n^C$ are geometrical objects on M^2. They determine $i_{\mu\nu}$ and $e_{\mu\nu}$ on a spherically symmetric space time M uniquely and they are assumed to be continuous across the contours of $f(x^C)$.

On any given contour Σ of f the intrinsic derivative is defined by

$$D_A N = i_A{}^B N_{,B}$$

for a scalar N. For a vector a_B which obeys $a_B n^B = 0$ this derivative is

$$D_A a_B = i_A{}^C i_B{}^D a_{C|D}$$

Furthermore, D is compatible with the metric i_{AB} of Σ:

$$D_C i_{AB} = 0$$

and it satisfies the Leibnitz rule for the derivative of a product. Consequently, D is the covariant derivative of Σ.

III. CALCULUS OF EVEN PARITY GAUGE INVARIANTS

The theory of relativistic perturbations on a spherically symmetric space time on occasion is characterized by a proliferation of symbols that makes a survey of its overall structure seem quite laborious. For spherically symmetric backgrounds this is true, more so of even that of odd parity perturbations. Expressing perturbations in terms of geometrical objects (scalar, vector, and tensor fields) on a two-dimensional space–time eases, of course, the labor quite substantially. Nevertheless, after the process of casting the perturbations (and the equations they satisfy) into an optimally efficient notation for the purpose of surveying their structure, there still remains the problem of preventing confusion between perturbations of different background geometrical objects. These objects are:

(i) the scalar f, one of whose contours is the submanifold Σ (say, the history of the surface of the collapsing star) across which the stress–energy tensor has a jump discontinuity,

(ii) the vector field of unit normals $n_\mu = (f_{,\alpha} f_{,\beta} g^{\alpha\beta})^{-1/2}$ on each of the contours of f,

(iii) the background metric tensor field $g_{\mu\nu}$,

(iv) the metric tensor field $i_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu$, intrinsic to each of the contours of f,

(v) the extrinsic curvature $e_{\mu\nu} = n_{\alpha;\beta} i_\mu{}^\alpha i_\nu{}^\beta = n_{\mu;\beta} i_\nu{}^\beta$ for each of the contours, and finally

(vi) the stress–energy tensor $t_{\mu\nu}$.

If one wanted to consider special stress energy tensors (e.g., fluid, electromagnetism, etc.) then there would be additional scalars and vectors to be kept track of.

With each of these objects are associated others, namely: a perturbation and a gauge invariant constructed from the perturbation.

To keep the number of new symbols at a minimum we have found the following notational rule successful: *Use lower case letters for the background geometrical objects; indicate their perturbations by the prefix Δ and their gauge invariants by using capitals.* We apply this rule consistently to all geometrical objects except to the metric tensor $g_{\mu\nu}$, where we attempt to deviate as little as possible from the now well-established notation of Regge and Wheeler. See Table I.

The construction of the gauge invariant geometrical perturbation objects in spherically symmetric space time is straightforward and runs parallel for scalars, vectors and tensors. It is accomplished by a four-step procedure:

1. Write down the background geometrical objects on a spherically symmetric space–time:

$$f = f(x^C), \quad \text{(scalar)}, \tag{3.1}$$

$$n_\mu dx^\mu = n_A(x^C) dx^A \quad \text{(vector)}, \tag{3.2}$$

$$t_{\mu\nu} dx^\mu dx^\nu = t_{AB}(x^C) dx^A dx^B + \tfrac{1}{2} t^c_{\ c}(x^C) r^2 \gamma_{ab} dx^a dx^b$$

(general symmetric tensor), (3.3)

TABLE I. Notation for even parity perturbational geometrical objects and their gauge invariants on a generic spherically symmetric space–time.

	Background geometrical object to M	Perturbation on M	Harmonic coefficients as geometrical objects on M^2				Gauge invariant geometrical perturbation objects on M^2			
			Y	$Y_{,a}$	$\gamma_{ab}Y$	$Y_{,a\,b}$	Tensor	Vector	1st scalar	2nd scalar
Metric tensor	$g_{\mu\nu}$	$h_{\mu\nu}$	h_{AB}	h_A	r^2K	r^2G	k_{AB}	0	k	0
Scalar	f	$\Delta f'$	Δf	—	—	—	—	—	F	—
Vector field	n_μ	Δn_μ	Δn_A	Δn	—	—	—	N_A	N	—
Intrinsic metric	$i_{\mu\nu}$	$\Delta i_{\mu\nu}$	Δi_{AB}	Δi_A	r^2K	r^2G	$k_{AB} - n_A N_B - n_B N_A$	$-n_A N$	k	0
Extrinsic curvature	$e_{\mu\nu}$	$\Delta e_{\mu\nu}$	Δe_{AB}	Δe_A	$r^2\Delta e^1$	Δe^2	E_{AB}	E_A	E^1	E^2
Matter tensor	$t_{\mu\nu}$	$\Delta t_{\mu\nu}$	Δt_{AB}	Δt_A	$r^2\Delta t^1$	Δt^2	T_{AB}	T_A	T^1	T^2

$$g_{\mu\nu}dx^\mu dx^\nu = g_{AB}(x^C)dx^A dx^B + r^2\gamma_{ab}dx^a dx^b \quad \text{(metric)}, \tag{3.4}$$

where

$$\gamma_{ab}dx^a dx^b = d\theta^2 + \sin^2\theta\, d\varphi^2$$

and $t_c{}^c = t_2{}^2 + t_3{}^3$ is the partial trace of $t_{\mu\nu}$.

2. Write down the perturbations (suppress angular integers l and m):

$$\Delta f' = \Delta f(x^C)Y(\theta,\varphi) \quad \text{(scalar)},$$

$$\Delta n_\mu dx^\mu = \Delta n_A(x^C)Y dx^A + \Delta n(x^C)Y_{,a}dx^a \quad \text{(vector)},$$

$$\begin{aligned}&\Delta t_{\mu\nu}dx^\mu dx^\nu \\ &= \Delta t_{AB}(x^C)Y dx^A dx^B + \Delta t_A(x^C)Y_{,a}(dx^A dx^a + dx^a dx^B) \\ &\quad + (r^2\Delta t^1 Y\gamma_{ab} + \Delta t^2 Y_{,a:b})dx^a dx^b \quad \text{(general symmetric tensor)},\end{aligned}$$

$$\begin{aligned}h_{\mu\nu}dx^\mu dx^\nu &= h_{AB}Y dx^A dx^B + h_A Y_{,a}(dx^A dx^a + dx^a dx^A) \\ &\quad + (r^2 K Y\gamma_{ab} + r^2 G Y_{,a:b})dx^a dx^b \quad \text{(metric)}.\end{aligned}$$

Here the coefficients of $Y, Y_{,a}, Y\gamma_{ab}$, and $Y_{,a:b}$ are geometrical objects on M^2.

3. Write down the "gauge" transformed geometrical perturbation objects. To do this, consider the even parity generator of an infinitesimal coordinate transformation

$$\xi_\mu dx^\mu = \xi_A Y dx^A + \xi Y_{,a}dx^a$$

and the associate Lie differentials of Eqs. (3.1)–(3.4):

$$\begin{aligned}&\mathcal{L}_\xi f = -f_{,\mu}\xi^\mu, \\ &\mathcal{L}_\xi n_\mu dx^\mu = -(n_{\mu;\sigma}\xi^\sigma + n_\sigma \xi^\sigma{}_{;\mu})dx^\mu, \\ &\mathcal{L}_\xi t_{\mu\nu}dx^\mu dx^\nu = -(t_{\mu\nu;\sigma}\xi^\sigma + t_{\sigma\nu}\xi^\sigma{}_{;\mu} \\ &\qquad + t_{\mu\sigma}\xi^\sigma{}_{;\nu})dx^\mu dx^\nu, \\ &\mathcal{L}_\xi g_{\mu\nu}dx^\mu dx^\nu = -(\xi_{\mu;\nu} + \xi_{\nu;\mu})dx^\mu dx^\nu.\end{aligned}$$

The "gauge" transformed geometrical perturbations objects are therefore

$$\bar{\Delta f} = \Delta f - f_{,B}\xi^B \quad \text{(scalar)}$$

$$\left.\begin{aligned}&\bar{\Delta n}_A = \Delta n_A - n_{A|B}\xi^B - n^B\xi_{B|A} \\ &\bar{\Delta n} = \Delta n - n_A\xi^A\end{aligned}\right\} \quad \text{(general vector)}$$

$$\left.\begin{aligned}&\bar{\Delta t}_{AB} = \Delta t_{AB} - t_{AB|C}\xi^C - t_{CB}\xi^C{}_{|A} \\ &\qquad - t_{AC}\xi^C{}_{|B} \\ &\bar{\Delta t}_A = \Delta t_A - t_{AC}\xi^C - \tfrac{1}{2}r^2 t_a{}^a(\xi/r^2)_{,A} \\ &\bar{\Delta t}^1 = \Delta t^1 - \tfrac{1}{2}r^{-2}(r^2 t_a{}^a)_{,A}\xi^A \\ &\bar{\Delta t}^2 = \Delta t^2 - t_a{}^a\xi\end{aligned}\right\} \quad \text{(general symmetric tensor)}$$

$$\left.\begin{aligned}&\bar{h}_{AB} = h_{AB} - \xi_{A|B} - \xi_{B|A} \\ &\bar{h}_A = h_A - \xi_A - r^2(\xi/r^2)_{,A} \\ &\bar{K} = K - 2v_A\xi^A \\ &\bar{G} = G - 2r^{-2}\xi \\ &\bar{p}_A = p_A - \xi_A\end{aligned}\right\} \quad \text{(metric)}$$

where

$$p_A \equiv h_A - \tfrac{1}{2}r^2 G_{,A} \tag{3.5}$$

4. Using the expression $\bar{G}$ and $\bar{p}_A$ derived from the metric, eliminate (by taking linear combinations) the "gauge" dependence in the other expressions, and obtain thereby the following gauge invariant geometrical perturbation objects:

$$f \to F = \Delta f - f_{,B}p^B \quad \text{(scalar)} \tag{3.6}$$

$$n_\mu \to \left.\begin{aligned}&N_A = \Delta n_A - n_{A|B}\,p^B - n_B\,p^B{}_{|A} \\ &N = \Delta n - n_B\,p^B\end{aligned}\right\} \quad \text{(general vector)}, \tag{3.7}$$

$$t_{\mu\nu} \to \left.\begin{aligned}&T_{AB} = \Delta t_{AB} - t_{AB|C}\,p^C \\ &\qquad - t_A{}^C p_{C|B} - t_B^C p_{C|A} \\ &T_A = \Delta t_A - t_{AC}\,p^C - \tfrac{1}{4}r^2 t_a{}^a G_{,A} \\ &T^1 = \Delta t^1 - \tfrac{1}{2}r^{-2}(r^2 t_a{}^a)_{,B}\,p^B \\ &T^2 = \Delta t^2 - \tfrac{1}{2}r^2 t_a{}^a G\end{aligned}\right\} \begin{matrix}\text{(general}\\ \text{symmetric}\\ \text{tensor),}\end{matrix} \tag{3.8}$$

$$g_{\mu\nu} \to \left.\begin{aligned}&k_{AB} = h_{AB} - p_{A|B} - p_{B|A} \\ &k = K - 2v_A\,p^A\end{aligned}\right\} \quad \text{(metric)} \tag{3.9}$$

where "$\to$" means "has the gauge invariant perturbation object(s)". It is clear that the symbol $\to$ is a linear operator: The gauge invariant of a given linear combination is the linear combination of the corresponding gauge invariants of each term of the given combination. There are three noteworthy features of the above collection of gauge invariants.

(i) Any given background geometrical object on four-dimensional space–time possesses a hierarchy of gauge invariant perturbational objects on M^2 with a tensor of rank zero (a scalar) at the bottom and a tensor of the same rank as that of the background object at the top of the hierarchy.

(ii) All gauge invariants depend on the metric perturbations vector p_A and scalar G. The hierarchy of metric gauge invariants has therefore fewer members; the vector and a scalar member are by necessity identically zero. See Table I.

(iii) Standard formulations of nonrelativistic as well as relativistic hydrodynamical perturbation theory draw a sharp distinction between Eulerian and Lagrangian perturbational variables. This is merely a difference in the choice of gauge. In gauge invariant perturbation theory such a distinction is superfluous, at least on a spherically symmetric background. Thus the gauge invariant perturbation set of the matter 4-velocity vector field v_μ, for example, is

$$v_\mu \to \left.\begin{aligned}&V_A = \Delta v_A - v_{A|B}\,p^B - v_B\,p^B{}_{|A}, \\ &V = \Delta v - v_B\,p^B.\end{aligned}\right.$$

The Lagrangian perturbational variables are recovered by working in the Lagrangian gauge which is obtained by setting to zero the spatial components of the velocity perturbation,

$$\Delta v_1 = 0, \quad \Delta v = 0.$$

Eulerian variables as such are not well defined because they can be altered at will by some infinitesimal coordinate transformation $\bar{x}^\mu = x^\mu + \xi^\mu$. This is the reason why workers in the field work mostly in the Lagrangian gauge. It is well defined and the perturbational variables coincide therefore with the corresponding gauge invariants. By introducing gauge invariant perturbation objects reference to Lagrangian or Eulerian variables can be eliminated from the very start.

We must now construct the perturbational gauge invariants for some special tensor fields, such as the first and second fundamental forms, $i_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu$ and $e_{\mu\nu} = n_{\alpha;\beta} i_\mu{}^\alpha i_\nu{}^\beta$ on the contours of the scalar f. Specifically, it is necessary to express these gauge invariants in terms of those associated with the unit normal n_μ to Σ and in terms of the metric gauge invariants k_{AB} and k. This is clearly necessary because the availability of such expressions gives the sought after junction conditions. Instead of becoming embroiled in a complex and tedious perturbation computation, we break up the task into a number of small problems and thereby obtain the answer in a systematic way. This we do in a sequence of identities with the property that each is obtained by using the preceeding ones only.

The task is simplest for scalars and vectors. Thus for any two vector fields u_α and v_β, for example, one has using Eqs. (3.6), (3.7), and (3.9),

$$v_\alpha u_\beta g^{\alpha\beta} \to V_B u^B + v_B U^B - v^A u^B k_{AB} \,. \tag{3.10}$$

Thus

$$n_\alpha n_\beta g^{\alpha\beta} = 1 \to N_B n^B - \tfrac{1}{2} n^A n^B k_{AB} = 0 \,. \tag{3.11}$$

This is a constraint that has to be satisfied by the perturbation of n_μ. For any two scalar fields h and f one has, using Eqs. (3.6)–(3.7),

$$n_\mu = hf_{,\mu} \to \begin{cases} N_A = hF_{,A} + Hf_{,A}, & \text{(3.12a)} \\ N = hF. & \text{(3.12b)} \end{cases}$$

Consider now a special case. If $n_\mu = (f_{,\alpha} f_{,\beta} g^{\alpha\beta})^{-1/2} \times f_{,\mu}$ is the tensor field of unit normals to the contours of f, then (3.10) and (3.12) imply

$$h = (f_{,\alpha} f_{,\beta} g^{\alpha\beta})^{-1/2} \to H = -h\,(hn^B F_{,B} - \tfrac{1}{2} n^C n^D k_{CD}) \,. \tag{3.13}$$

Furthermore, using the above definitions of h and n_μ

$$h_{,B} i^B{}_A = -hn_{A|C} n^C \equiv -ha_A \,,$$

where a_A is the acceleration of an observer whose world line is normal to (the contours of) f. We shall see below that the quantity continuous across these contours is N, not F. Thus we express Eq. (3.12a) in terms of N. Using Eqs. (3.12b) and (3.13), one obtains

$$N_A = (N_{,B} + n_{B|C} n^C N) i_A{}^B + \tfrac{1}{2} N^C n^B k_{CB} n_A \tag{3.14}$$

in terms of the surface deformation scalar N.

For tensors the task is analogous to that used to obtain Eqs. (3.12). Thus, having obtained the gauge invariant object(s) associated with some scalar, vector or tensor, reexpress the object(s) in terms of gauge invariants that have been constructed already. The four gauge invariant objects of the symmetric tensor $t_{\mu\nu} = f_{,\mu;\nu}$, for example, are constructed using Eqs. (3.8). The terms of each of these four expressions are regrouped into the gauge invariants F and k_{AB} and their derivatives. The result is

$$t_{\mu\nu} = f_{,\mu;\nu} \to \begin{cases} T_{AB} = F_{,A|B} - \tfrac{1}{2} f^{|C} \nabla_C k_{AB}, \\ T_A = F_{,A} - Fv_A - \tfrac{1}{2} f^{|C} k_{CA}, \\ T^1 = v^C F_{,C} - f^{|B} v^C k_{BC} + \tfrac{1}{2} r^{-2} f^{|C} (r^2 k)_{|C}, \\ T^2 = F, \end{cases} \tag{3.15}$$

where we write $\nabla_C k_{AB} \equiv k_{CA|B} + k_{CB|A} - k_{AB|C}$, and v_C is defined by Eq. (2.2). The reduced form of the perturbations of the derivatives of $h_{\mu\nu}$, which occur in $\Delta\Gamma_{\mu\nu}{}^\sigma = \tfrac{1}{2}(h_\mu{}^\sigma{}_{;\nu} + h_\nu{}^\sigma{}_{;\mu} - h_{\mu\nu}{}^{;\sigma})$, are not listed here. Other gauge invariants are obtained in the same way. Thus for any given background scalar h and symmetric tensor $t_{\mu\nu}$ one has

$$ht_{\mu\nu} \to \begin{cases} hT_{AB} + t_{AB} H, \\ hT_A, \\ hT^1 + \tfrac{1}{2} t_a{}^a H, \\ hT^2. \end{cases} \tag{3.16}$$

If $t_{\mu\nu}$ is the symmetrized product of any two given vector fields u_μ and w_ν then

$$t_{\mu\nu} = u_{(\mu} w_{\nu)} \to \begin{cases} T_{AB} = \tfrac{1}{2}(u_A W_B + u_B W_A) \\ \qquad\quad + \tfrac{1}{2}(U_A w_B + U_B w_A), \\ T_A = \tfrac{1}{2}(u_A W + U w_A), \\ T^1 = 0, \\ T^2 = 0. \end{cases} \tag{3.17}$$

The linearity of "$\to$" implies that the gauge invariants associated with $n_{(\mu;\nu)} = \tfrac{1}{2}[(hf_{,\mu})_{;\nu} + (hf_{,\nu})_{;\mu}]$ are, with the help of Eqs. (3.15)–(3.17),

$$n_{(\mu;\nu)} \to \begin{cases} \tfrac{1}{2}(hF_{,A} + f_{,A} H)_{|B} + (A \leftrightarrow B) - (n^C/2) \nabla_C k_{AB}, \\ h\,(F_{,A} - Fv_A - \tfrac{1}{2} f^{|C} k_{CA}) + \tfrac{1}{2}(h_{,A} F + Hf_{,A}), \\ h\,(v^C F_{,C} - f^{|B} v^C k_{BC}) + \tfrac{1}{2} r^{-2} hf^{|C} (r^2 k)_C + v^C f_{,C} H, \\ hF \end{cases}$$

or with Eqs. (3.12)

$$n_{(\mu;\nu)} \to \begin{cases} N_{(A|B)} - \tfrac{1}{2} n^C \nabla_C k_{AB}, \\ \tfrac{1}{2} N_A + \tfrac{1}{2} N_{,A} - Nv_A - \tfrac{1}{2} n^C k_{CA}, \\ v^C (N_C - n^B k_{BC}) + \tfrac{1}{2} r^{-2} (r^2 k)_{|C} n^C, \\ N. \end{cases} \tag{3.18}$$

The gauge invariant perturbation of the intrinsic metric are obtained using (3.9), (3.17) and the linearity of "$\to$":

$$i_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu \to \begin{cases} I_{AB} = k_{AB} - n_A N_B - n_B N_A, \\ I_A = -n_A N, \\ I^1 = k, \\ I^2 = 0. \end{cases} \tag{3.19}$$

Given any two symmetric tensor fields $t_{(\mu\nu)}$ and $i_{\alpha\beta}$, the gauge invariant perturbation of $t_{(\alpha\beta)} i^\alpha{}_\mu i^\beta{}_\nu = e_{(\mu\nu)}$ is

$$e_{(\mu\nu)} \to \begin{cases} E_{AB} = i_A{}^C i_B{}^D T_{(CD)} + t_C{}^D i_A{}^C I_{DB} + t_C{}^D i_B{}^C I_{DA} \\ \qquad - t_E{}^C i_A{}^E i_B{}^D k_{CD} - t_E{}^C i_B{}^E i_A{}^D k_{CD}, \\ E_A = \tfrac{1}{2} i_a{}^a i_A{}^C T^C + \tfrac{1}{4} i_a{}^a t_b{}^b I_A + i_A{}^C t_C{}^D I_D, \\ E^1 = \tfrac{1}{2} i_a{}^a i_c{}^c T^1 + \tfrac{1}{4} t_a{}^a i_c{}^c I^1 - \tfrac{1}{8} t_a{}^a i_b{}^b i_c{}^c k, \\ E^2 = \tfrac{1}{4} i_a{}^a i_b{}^b T^2 + \tfrac{1}{4} t_a{}^a i_b{}^b I^2. \end{cases}$$

If the tensor $i_{\mu\nu}$ is the intrinsic metric given in Eq. (3.19), then the gauge invariant perturbations of $e_{(\mu\nu)}$ reduce to

$$e_{(\mu\nu)} \to \begin{cases} E_{AB} = i_A{}^C i_B{}^D T_{(CD)} - 2n^B t_{BC} i^C{}_{(A} N_{B)} \\ \qquad + 2(n^D k_{DE} - N_E) t_C{}^E i^C{}_{(A} n_{B)}, \\ E_A = (T_C - n_D t_C{}^D N) i_A{}^C - \tfrac{1}{2} t_a{}^a n_A N, \\ E^1 = T^1, \\ E^2 = T^2. \end{cases} \tag{3.20}$$

Use Eq. (3.20) to obtain the gauge invariant perturbation objects of $e_{(\mu\nu)} = n_{(\alpha;\beta)} i^{\alpha}{}_{\mu} i^{\beta}{}_{\nu}$. They are given with the help of (3.18), (3.12) and

$$\tfrac{1}{2} t_a{}^a = \tfrac{1}{2} n_a{}^{;a} = v_C n^C, \; t_{BC} = \tfrac{1}{2}(n_{B|C} + n_{C|B})$$

as follows

$$e_{\mu\nu} = n_{(\alpha;\beta)} i^{\alpha}{}_{\mu} i^{\beta}{}_{\nu} \rightarrow$$

$$\rightarrow \begin{cases} E_{AB} = [N_{C|D} - \tfrac{1}{2} n^E v_E k_{CD}] i_A{}^C i_B{}^D - n^D n_{C|D} i^C{}_{(A} N_{B)} \\ \qquad + (n^D k_{DE} - N_E)(n_{|C}{}^E + n_C{}^{|E}) i^C{}_{(A} n_{B)}, & (3.21a) \\ E_B = [N_{,B} - N v_B - \tfrac{1}{2} n^D k_{DB}] i_A{}^B - v_B n^B n_A N, & (3.21b) \\ E^1 = v^B (N_B - n^C k_{CB}) + \tfrac{1}{2} r^{-2} (r^2 k)_{|C} n^C, & (3.21c) \\ E^2 = N, & 3.21d) \end{cases}$$

$$E_{AB} i^{AB} = D_A D^A N + (D_C a^C - a_C a^C) N + \tfrac{1}{2} n^C n^B k_{CB} n_A{}^{|A} - \tfrac{1}{2} n^E \nabla_E k_{AB} i^{AB}, \quad (3.22)$$

where

$$a_C = n_{C|B} n^B$$

is the acceleration of an observer on Σ and

$$\nabla_E k_{AB} = k_{EA|B} + k_{EB|A} - k_{AB|E} .$$

Equations (3.19) and (3.21) are the respective sets of the gauge invariant perturbation objects of the first and second fundamental forms. In the next section we obtain those linear combinations of these expressions that are continuous across any given contour of Σ.

IV. CONTINUOUS EVEN PARITY GAUGE INVARIANTS CONSTRUCTED FROM INTRINSIC METRIC

First obtain the junction conditions obtained from the perturbations of the intrinsic metric field $i_{\mu\nu}$; then in the next section those associated with the extrinsic curvature tensor $e_{\mu\nu}$.

An even parity perturbation in the metric

$$i_{\mu\nu} dx^\mu dx^\nu = (g_{AB} - n_A n_B) dx^A dx^B + r^2 \gamma_{ab} dx^a dx^b$$

intrinsic to any one of the contours of f, which have spacelike normals n_μ, is

$$\Delta i_{\mu\nu} dx^\mu dx^\nu = \Delta i_{AB} Y dx^A dx^B + \Delta i_A Y_{,a} (dx^A dx^a + dx^a dx^A) + (r^2 K Y \gamma_{ab} + r^2 G Y_{,a;b}) dx^a dx^b) . \quad (4.1)$$

The continuous perturbation objects are the projections of $\Delta i_{\mu\nu}$ onto the contours of f.[10] Thus the following four objects are continuous:

$$i_A{}^C i_B{}^D \Delta i_{CD} = h_{CD} i_A{}^C i_B{}^D, \qquad i_A{}^C \Delta i_C = i_A{}^C h_C , \quad K, \text{ and } r^2 G \quad (4.2)$$

In addition

$$D_A (r^2 G) \equiv i_A{}^C (r^2 G)_{,C} \quad (4.3)$$

is also continuous because it is the intrinsic derivative of $r^2 G$. It follows that the intrinsic vector

$$i_B{}^C p_C \equiv i_B{}^C [h_C - \tfrac{1}{2}(r^2 G)_{,C}] \quad (4.4a)$$

as well as its intrinsic derivative

$$D_A (i_B{}^C p_C) \equiv i_A{}^E i_B{}^D (i_D{}^C p_C)_{|E} \quad (4.4b)$$

are continuous.

Furthermore, it follows from Eqs. (4.2) and (4.4) that

$$k = K - 2 v_B i_C{}^B p_C - 2 v_B n^B n_C p^C \quad (4.5a)$$

$$k_{CD} i_A{}^C i_B{}^D = h_{CD} i_A{}^C i_B{}^D - D_A (i_B{}^C p_C) - D_B (i_A{}^C p_C) - 2 n_{D|E} i_A{}^D i_B{}^E n_C p^C , \quad (4.5b)$$

obtained from Eqs. (1.4), have each of their terms, except their last ones, continuous. A single continuous gauge invariant is obtained from Eqs. (4.5a), (4.5b) by eliminating these offending terms between these two equations. This can be done because the coefficients ($2 v_B n^B = e_a{}^a$ and and $2 e_{AB}$) of $n_C p^C$ are continuous. The result of this elimination is the continuous gauge invariant scalar

$$e_B{}^B k - v_B n^B k_{CD} i^{CD} = n_B{}^{|B} k - v_B n^B (k_C{}^C - k_{CD} n^C n^D). \quad (4.6)$$

Here we used

$$k_{CD} i_A{}^C i_B{}^D = k_{CD} i^{CD} i_{AB} ,$$

which always holds in two dimensions if the symmetric tensor field is intrinsic to the contours of f.

V. CONTINUOUS GAUGE INVARIANTS CONSTRUCTED FROM EXTRINSIC CURVATURE.

The remaining continuous gauge invariants are obtained from the continuous extrinsic curvature

$$e_{\mu\nu} = n_{\alpha;\beta} i_\mu{}^\alpha i^\beta{}_\nu \quad (5.1)$$

Its perturbation (supress angular integer l and m) is

$$\Delta e_{\mu\nu} dx^\mu dx^\nu = \Delta e_{AB} Y dx^A dx^B + \Delta e_A Y_{,a} (dx^A dx^a + dx^a dx^A) + (r^2 \Delta e^1 Y \gamma_{ab} + \Delta e^2 Y_{,a;b}) dx^a dx^b . \quad (5.2)$$

The continuous perturbation objects are the projections of $\Delta e_{\mu\nu}$ onto the contours of f.[10] Thus the following four expressions are continuous:

$$\Delta e_{CD} i_A{}^C i_B{}^D = \Delta e_{CD} i^{CD} i_{AB} , \quad (5.3a)$$

$$i_A{}^C \Delta e_C , \quad (5.3b)$$

$$\Delta e^1, \text{ and } \Delta e^2 . \quad (5.3c,d)$$

These four objects are, however, not gauge invariant. The four corresponding objects that are gauge invariant are obtained by referring to Eqs. (3.8) and by projecting, whenever appropriate, the gauge invariant objects onto the contours of f. The result is

$$i^{AB} E_{AB} = i^{AB} [\Delta e_{AB} - e_{AB|C} p^C - e_{AC} p_{|B}{}^C - e_{BC} p_{|A}{}^C] , \quad (5.4a)$$

$$i_A{}^B E_B = i_A{}^B [\Delta e_B - \tfrac{1}{4} r^2 e_a{}^a G_{,B} - e_{BC} p^C] , \quad (5.4b)$$

$$E^1 = \Delta e^1 - \tfrac{1}{2} r^{-2} (r^2 e_a{}^a)_{|B} p^B , \quad (5.4c)$$

$$E^2 = \Delta e^2 - \tfrac{1}{2} r^2 e_a{}^a G . \quad (5.4d)$$

Here Eq. (5.4a) is the trace of

$$E_{CD} i_A{}^C i_B{}^D = E_{CD} i^{CD} i_{AB} .$$

Although Eqs. (5.4a)–(5.4d) are not all continuous across the contours of f, they at least are *nearly* continuous. Indeed one sees from Eqs. (5.3d) and (4.2) that Eq. (5.4d) is continuous. Similarly (5.4b) is also continuous because

$e_{BC}\,p^C = e_D{}^D i_{BC}\,p^C$ is continuous according to Eq. (4.4a).

The properties of e_{AB} and i_{AB} imply that the remaining gauge invariants (5.4a) and (5.4c) are

$$E_{CD}i^{CD} = \Delta e_{CD}i^{CD} - e_B{}^B{}_{|D}i^D{}_C\,p^C{}_B - 2e_B{}^B(i_C{}^D p_D)_{|E}i^{CE} - (e_B{}^B{}_{|D}n^D + 2e^{BD}n_{B|D})n_C\,p^C,$$

$$E^1 = \Delta e^1 - \tfrac{1}{2}r^{-2}(r^2 e_a{}^a)_{|B}i^B{}_C\,p^C - (e_a{}^a v_B n^B + \tfrac{1}{2}e_a{}^a{}_{|B}n^B)n_C\,p^C.$$

All terms except the last ones containing the factor $n_C\,p^C$ are continuous. Indeed,

$$E_{CD}i^{CD} + 2E^1 = (\text{continuous terms}) - [2e^{BD}n_{B|D} + 2e_a{}^a v_B n^B + (e_B{}^B + e_a{}^a)_{|D}n^D]n_C\,p^C \tag{5.5}$$

has the contents of its square brackets continuous. This follows from the fact that $(e_B{}^B + e_a{}^a)_{|D}n^D$ is continuous. By eliminating the last discontinuous terms between Eq. (5.5) and Eq. (4.5a) one obtains the last of the three continuous gauge invariant obtained from the extrinsic curvature:

$$2(E_{AB}i^{AB} + 2E^1)v_C n^C - k\,[2e^{BD}n_{B|D} + 2e_a{}^a v_B n^B + (e_B{}^B + e_a{}^a)_{|D}n^D] \tag{5.6}$$

We have shown therefore that this expression together with $E_B i_A{}^B$ and E^2 as well as Eq. (4.6) are the four even parity continuous gauge invariant perturbation objects. It remains to actually exhibit these expressions in terms of the familiar gauge invariant metric perturbations k, k_{AB}, and the gauge invariant surface perturbation scalar N.

To this end, substitute Eq. (3.21d) for Eq. (5.4d), Eq. (3.21b) for (5.4b), introduce Eqs. (3.21c) and (3.22) into Eq. (5.6), simplify, and obtain three continuous even parity gauge invariant perturbation objects from the extrinsic curvature:

$$E^2 = N, \tag{5.7a}$$

$$i_A{}^B E_B = [N_{,B} - v_B\dot{N} - \tfrac{1}{2}n^C k_{CB}]i_A{}^B, \tag{5.7b}$$

$$(5.6) = 2n^C v_C[D_A D^A N + 2v^A D_A\dot{N} + (D_A a^A - a_A a^A + 2v_A a^A)N + n^C n^D k_{CD}(v_A n^A + \tfrac{1}{2}n_A{}^{|A}) - \tfrac{1}{2}n^E\nabla_E k_{AB}i^{AB}] + 2n_C v^C n^A k_{,A} - \beta k, \tag{5.7c}$$

where

$$\beta = 2n_C{}^{|C}n_D{}^{|D} + (n_C{}^{|C} + 2n_C v^C)_{|D}n^D.$$

The continuity requirements (5.7b) and (5.7c) are simplified by observing that the continuity of (5.7a) implies the continuity of its intrinsic derivatives

$$D_B N \equiv N_{,E}i_B{}^E, \tag{5.8a}$$

$$D_A D_B N \equiv [N_{,E}i_C{}^E]_{|D}i_B{}^C i_A{}^D. \tag{5.8b}$$

These continuous expressions together with the continuous scalars $e_A{}^A = n_A{}^{|A}$ and $e_a{}^a = 2v_A n^A$ can be used to obtain a different but equivalent set of continuous objects. Thus, with help of Eqs. (5.7a) and (5.8a), the continuity of Eq. (5.7b) can be replaced by the requirement that

$$n^C k_{CB}i_A{}^B \tag{5.7b'}$$

be continuous.

Similarly, by subtracting obvious multiples of Eqs. (5.8a) and (5.8b) from Eq. (5.7e), the remaining continuous expression is brought into a form having no derivatives of N. The result is given by Eq. (1.8d).

Equations (5.7a), (5.7b'), (5.7c) together with Eq. (4.6), or equivalently Eqs. (1.8a)–(1.8d) are the four perturbational gauge invariants that are continuous across the timelike interface Σ, across which the background first and second fundamental form are continuous. The deformations of Σ are determined by N_A, Eqs. (3.14) or (1.9).

CONCLUSION

Suppose one wishes to identify and classify perturbations of a spherically symmetric, in general matter-occupied, space-time. The most natural way of accomplishing this is in terms of geometrical objects that arise from, say, the linerized Einstein field equations. However, without specifying the background geometry these equations and its metric perturbation tensor are of little practical use. Furthermore, as is well known, for a given perturbation, this tensor description is nonunique, reflecting the fact that perturbations in the background coordinate system (i.e. gauge transformations) yield different tensors.

On the other hand, if one does specify the background geometry by having the background metric expressed with respect to a specific spherical coordinate system, then one has excluded other possibly more appropriate and natural coordinate systems from the very start. One should not have to make a coordinate commitment until the very last moment of one's perturbational analysis. In addition, if one does force uniqueness onto the perturbation tensor description by fixing the gauge representation of the tensor, then one has injected an unwarranted and probably an undesirable arbitrariness into the representation of the perturbation.

The description of perturbations in terms of gauge invariant geometrical objects defined on M^2 avoids all four disadvantages listed above; it captures the best of both worlds: namely, the geometric formulation (i.e., no coordinate commitment) and the gauge invariance (i.e., uniqueness of the representation). This optimal formulation applies not only to the linearized Einstein field equations[5] but also to the junction conditions both for odd parity[10] and for even parity perturbations as we have seen in this paper.

[1]See, e.g., C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 21.
[2]See, e.g., K. Thorne, in *Theoretical Principles in Astrophysics and Relativity*, edited by N. Lebovitz *et al.* (Univ. of Chicago Press, Chicago, 1978).
[3]S.A. Colgate and R.H. White, Astrophys. J. **143**, 626 (1966).
[4]M.M. May and R.H. White, Phys. Rev. **141**, 1232 (1966).
[5]U.H. Gerlach and U.K. Sengupta, Phys. Rev. D **19**, 2268 (1979).
[6]D. Kazanas, Astrophys. J. Lett. **22**, L109 (1978).
[7]D. Kanzanas and D.N. Schramm, Astrophys. J. **214**, 819 (1977).
[8]In the sense first introduced by V.E. Moncrief, Ann. Phys. **88**, 323 (1974).
[9]This expression corrects a misplaced parenthesis in Eq. (4b) of Ref. 5.
[10]U.H. Gerlach and U.K. Sengupta, Phys. Rev. D **19**, 2268 (1979).

An extension of the Bonnor transformation[a)]

Elliot Fischer[b)]
Systems Evaluation Division, Institute for Defense Analyses, 400 Army–Navy Drive, Arlington, Virginia 22202

(Received 1 February 1979; accepted for publication 8 May 1979)

The Bonnor transformation, which maps stationary, axially symmetric vacuum fields into static, axially symmetric Einstein–Maxwell fields is extended to spacetimes possessing only one spacelike Killing vector. The transformation generates two distinct Einstein–Maxwell solutions from any vacuum solution.

1. INTRODUCTION

Transformations that generate new solutions to the Einstein or Einstein–Maxwell equations from known ones have become an effective tool for producing new solutions and probing relationships between old ones. The Ehlers transformation[1] and the Bonnor transformation[2] are two well-known transformations for spacetimes possessing two commuting Killing vectors. Extensions of these transformations to spacetimes possessing one Killing vector are of great interest due to the complexity of the field equations and the resulting scarcity of exact solutions. The Ehlers transformation was extended by Kinnersley[3] in 1973. In this paper we present an extension of the Bonnor transformation.

The main idea of the Bonnor transformation is to map stationary vacuum fields into static Einstein–Maxwell ones. In Sec. 2 we review the general metric and the Einstein–Maxwell equations for spacetimes with one Killing vector. In Sec. 3 we consider some particular cases and explicitly show a formal analogy between the vacuum Einstein equations and the Einstein–Maxwell equations that may be used to generate new, physically distinct solutions from any known solution. This analogy is the extension of the original Bonnor transformation to more general spacetimes.

2. ASSUMED METRIC AND ITS EINSTEIN–MAXWELL EQUATIONS

A detailed account of the derivation of the Einstein–Maxwell equations with one Killing vector has previously been presented by Harrison.[4] We outline the important parts of the derivation as follows.

Sign and other conventions are as in Landau and Lifshitz.[5] The metric is assumed to have the form

$$-ds^2 = \epsilon e^{2U}(dx^k + a f_\alpha dx^\alpha)^2 + a^2 e^{-2U}\gamma_{\alpha\beta}dx^\alpha dx^\beta, \tag{1}$$

where a is an arbitrary constant, $\epsilon = \pm 1 = \text{sign}\,(g_{kk})$. k is some one particular value of 0,1,2,3. Greek letters take all values of 0,1,2,3 except k, and all metric coefficients are independent of x^k. Therefore, a Killing vector field is generated by translation along x^k. Latin letters (except k) take on all values of 0,1,2,3 except k, and all metric coefficients are independent of x^k. Therefore, a Killing vector field is generated by translation along x^k. Latin letters (except k) take on all values of 0,1,2,3. If $\epsilon = -1$, the Killing vector is timelike and $k = 0$; if $\epsilon = +1$, it is spacelike.

For notational convenience, we define the differential parameters

$$\begin{aligned}\Delta_1(F) &= \gamma^{\alpha\beta}F_{,\alpha}F_{,\beta},\\ \Delta_1(F,G) &= \gamma^{\alpha\beta}F_{,\alpha}G_{,\beta},\\ \Delta_2(F) &= \gamma^{\alpha\beta}F_{;\alpha\beta},\end{aligned} \tag{2}$$

where $\gamma_{\alpha\beta}$ is the inverse of the three-dimensional metric $\gamma_{\alpha\beta}\cdot\gamma_{\alpha\beta}$ may be interpreted as a background 3-metric in the 3-space that is the quotient of the 4-space by the Killing vector. A comma denotes partial differentiation with respect to x^α while a semicolon denotes covariant differentiation with respect to $\gamma_{\alpha\beta}$.

The Einstein–Maxwell equations for the metric (1) can be written as[4]

$$\begin{aligned}&\Delta_2 + \epsilon e^{-2U}[\Delta_1(A) + \Delta_1(B)] + \tfrac{1}{2}e^{-4U}[\Delta_1(\Phi)\\ &\quad + 4\epsilon B\Delta_1(\Phi,A) - 4\epsilon A\Delta_1(\Phi,B) + 4B^2\Delta_1(A)\\ &\quad - 8AB\Delta_1(A,B) + 4A^2\Delta_1(B)] = 0,\end{aligned} \tag{3}$$

$$\begin{aligned}&\Delta_2(\Phi) - 4\Delta_1(U,\Phi) + 2\epsilon B[\Delta_2(A) - 4\Delta_1(U,A)]\\ &\quad - 2\epsilon A[\Delta_2(B) - 4\Delta_1(U,B)] = 0,\end{aligned} \tag{4}$$

$$\begin{aligned}&\Delta_2(A) - 2\Delta_1(U,A) + e^{-2U}[\Delta_1(\Phi,B) + 2\epsilon B\Delta_1(A,B)\\ &\quad - 2\epsilon A\Delta_1(B)] = 0,\end{aligned} \tag{5}$$

$$\begin{aligned}&\Delta_2(B) - 2\Delta_1(U,B) - e^{-2U}[\Delta_1(\Phi,A) + 2\epsilon B\Delta_1(A)\\ &\quad - 2\epsilon A\Delta_1(A,B)] = 0,\end{aligned} \tag{6}$$

$$\begin{aligned}P_{\alpha\beta} &= 2U_{,\alpha}U_{,\beta} + 2\epsilon e^{-2U}(A_{,\alpha}A_{,\beta} + B_{,\alpha}B_{,\beta})\\ &\quad + \tfrac{1}{2}e^{-4U}[\Phi_{,\alpha} + 2\epsilon(BA_{,\alpha} - AB_{,\alpha})]\\ &\quad \times[\Phi_{,\beta} + 2\epsilon(BA_{,\beta} - AB_{,\beta})].\end{aligned} \tag{7}$$

$P_{\alpha\beta}$ is the three-dimensional Ricci tensor for the background metric $\gamma_{\alpha\beta}$.

A and B are electromagnetic potentials in terms of which the electromagnetic field tensor may be expressed as

$$\begin{aligned}F^{\alpha\beta} &= (-g)^{-1/2}\epsilon^{\alpha\beta\gamma}A_{,\gamma},\\ F_{k\alpha} &= B_{,\alpha}.\end{aligned}$$

Φ is a "twist" potential from which the f_α may be recovered, as follows. Define

$$h_{\alpha\beta} = f_{\alpha,\beta} - f_{\beta,\alpha}.$$

[a)]Research supported in part by NSF Grant MCS75-05415.

[b)]Work based on a thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy to the California Institute of Technology in May, 1977.

 0022-2488/79/122547-02$01.00

The $h_{\alpha\beta}$ has a dual axial vector:

$$h_{\alpha\beta} = \epsilon_{\alpha\beta\gamma}\gamma^{\gamma\delta}Z_\delta(-\gamma)^{1/2}.$$

Z_α is expressed in terms of U, A, B, and the potential Φ by

$$Z_\alpha = e^{-4U}[\Phi_{,\alpha} + 2\epsilon(BA_{,\alpha} - AB_{,\alpha})].$$

Φ is very similar to the twist potential for spacetimes with two Killing vectors. In fact, if $A = B = 0$ corresponding to pure vacuum solutions, then

$$h_{\alpha\beta} = \epsilon_{\alpha\beta\gamma}\gamma^{\gamma\delta}(-\gamma)^{1/2}e^{-4U}\Phi_{,\delta}$$

and Φ is seen to be the extension to three variables of the above mentioned twist potential. Furthermore, if $\Phi = 0$, then the f_α are all constants which can be set to zero and the Killing vector then defines an orthogonal congruence. If the Killing vector is timelike, the metric is then stationary.

3. PARTICULAR CASES

The use of a twist potential plays a central role in the original Bonnor transformation. The occurence of a similar twist potential here motivates us to look for a similar phenomena in Eqs. (3)–(7).

From the Einstein–Maxwell equations (3)–(7) we may obtain three important special (but physically distinct) cases, as follows:

Case I: $A = B = 0$

This is the pure vacuum case. Equations (3)–(7) reduce to

$$\Delta_2(\Psi) = -e^{-2\Psi}\Delta_1(\Phi), \tag{8}$$

$$\Delta_2(\Phi) = 2\Delta_1(\Psi,\Phi), \tag{9}$$

$$P_{\alpha\beta} = \frac{\Psi_{,\alpha}\Psi_{,\beta}}{2} + \tfrac{1}{2}e^{-2\Psi}\Phi_{,\alpha}\Phi_{,\beta}, \tag{10}$$

where $\Psi = 2U$.

Case II: $\Phi = A = 0$

The metric in the case $\Phi = 0$ has no terms of the form $dx^k dx^\alpha$ since $\Phi = 0$ implies $f_\alpha = 0$. Choosing $A = 0$ specifies a particular type of electromagnetic field. Equations (3)–(7) reduce to

$$\Delta_2(U) = -\epsilon e^{-2U}\Delta_1(B), \tag{11}$$

$$\Delta_2(B) = 2\Delta_1(U,B), \tag{12}$$

$$P_{\alpha\beta} = 2U_{,\alpha}U_{,\beta} + 2\epsilon e^{-2U}B_{,\alpha}B_{,\beta}. \tag{13}$$

Case III: $\Phi = B = 0$

The metric is the same as in Case II. The electromagnetic field is physically distinct from that of Case II. Equations (3)–(7) reduce to

$$\Delta_2(U) = -\epsilon e^{-2U}\Delta_1(A), \tag{14}$$

$$\Delta_2(A) = 2\Delta_1(U,A), \tag{15}$$

$$P_{\alpha\beta} = 2U_{,\alpha}U_{,\beta} + 2\epsilon e^{-2U}A_{,\alpha}A_{,\beta}. \tag{16}$$

It is immediately obvious that Eqs. (11)–(13) for Case II and Eqs. (14)–(16) for Case III are identical if we identify A and B. Furthermore, if we define new independent coordinates (denoted by primes) by

$$x'^\alpha = \tfrac{1}{2}x^\alpha, \tag{17}$$

but leave the background metric $\gamma_{\alpha\beta}$ untransformed, the equations for Cases II and III become, in terms of the primed variables [using Eq. (2) and omitting the primes]:

$$\Delta_2(U) = -\epsilon e^{-2U}\Delta_1(\Omega), \tag{18}$$

$$\Delta_2(\Omega) = 2\Delta_1(U,\Omega), \tag{19}$$

$$P_{\alpha\beta} = \frac{U_{,\alpha}U_{,\beta}}{2} + \frac{\epsilon e^{-2U}\Omega_{,\alpha}\Omega_{,\beta}}{2}, \tag{20}$$

where Ω is equal to either A or B.

Equations (18)–(20) for (U,Ω) are identical to Eqs. (8)–(10) for (Ψ,Φ) when $\epsilon = +1$. Thus, given any vacuum solution (Case I) we can generate two physically distinct electromagnetic solutions (when the Killing vector is spacelike) by using the transformation, Eq. (17). This can be considered an extension of the Bonnor transformation obtained for the stationary axisymmetric Einstein–Maxwell field equations with two commuting Killing vector fields, one spacelike and one timelike.[2] Since three-variable solutions of the Einstein and Einstein–Maxwell equations are relatively rare, the above transformation, which generates two new Einstein–Maxwell solutions for any given vacuum solution, should be very valuable.

ACKNOWLEDGMENTS

The author expresses appreciation to Dr. Frank B. Estabrook and Professor B. Kent Harrison for many conversations about the results of this paper.

[1] J. Ehlers, Dissertation, Hamburg, 1957.
[2] W. Bonnor, Z. Phys. **190**, 444 (1966); Z. Perjes, Nuovo Cimento B **55**, 600 (1968); R. Misra *et al.*, Phys. Rev. D 7, 2587 (1973).
[3] W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[4] B.K. Harrison, J. Math. Phys. **9**, 2744 (1968).
[5] L. Landau and E. Lifschitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, Mass., 1962).

Scalar quantum field in an external gravitational field

J. Dimock[a)]
Department of Mathematics, SUNY at Buffalo, Buffalo, New York 14214

(Received 16 April 1979; accepted for publication 25 June 1979)

We consider a class of Lorentz metrics on $\mathbb{R}^4$ which are Minkowskian off a compact set. For each such metric we construct a quantum field operator which satisfies the generalization of the Klein–Gordon equation. The field has a causal commutator and transforms as a scalar under general coordinate transformations. The field determines a unitary scattering operator which is invariant under coordinate transformations.

I. INTRODUCTION

This paper is concerned with the problem of a scalar quantum field in an external gravitational field. Alternatively, it is a problem of quantum field theory in curved space–time. The external field is specified by a Lorentz metric on the space–time manifold, and if $\Box$ is the d'Alembertian for the metric we ask for a quantum field operator φ satisfying the equation $(\Box + m^2)\varphi = 0$ where $m \geqslant 0$ is a mass parameter.[1]

This problem has attracted considerable interest lately due to a paper of Hawking[2] in which the metric is taken to be that of a collapsing black hole, and the prediction is made that the ensuing particle creation is sufficient to radiate everything away. This indicates a range of potential astrophysical applications. The external field problem may also be of interest as a building block in a more complete theory in which the metric is treated quantum mechanically as well. For a survey of what is known about the physics of quantum field theory in curved space–time, see DeWitt[3] or Parker.[4]

One of the goals of this paper is to give a mathematically precise construction of the field operator φ and its associated scattering theory for a certain nice class of metrics. The space–time manifold is taken to be $\mathbb{R}^4$ and the metric is required to be globally hyperbolic and Minkowskian off a compact set.[5] Under this assumption we show that there is a unique field operator φ solving $(\Box + m^2)\varphi = 0$ and reducing to a free field φ_{in} in the distant past. The field is causal in the sense that the commutator vanishes when the points are not causally related. The field reduces to a free field φ_{out} in the distant future and we show that there is a unitary scattering operator S satisfying $\varphi_{\text{out}} = S^{-1}\varphi_{\text{in}}S$. The existence of S uses an additional assumption on the metric which, however, is possibly no more restrictive.

Our approach to the external field problem follows particularly the general framework of Wightman.[6,7] In this framework the advanced and retarded fundamental solutions of the differential operator play a key role. In particular, estimates on the regularity of the kernel of the fundamental solutions are the basic input for the existence of the scattering operator. For these estimates we rely on the theory of Fourier integral operators as developed by Hörmander and Duistermaat.[8–10]

The existence of the scattering operator for this problem has been independently obtained by Wald in a recent preprint[11]; see also Ref. 12. The method are quite different from ours, and both approaches seem to be of interest. We also remark that progress has recently been made in obtaining a unitary scattering operator for other external field problems (e.g., electromagnetic) by Seiler,[13] Ruijsenaars,[14] and Dimock.[15]

A second goal of this paper is to show that the quantum dynamics we have defined is natural in the sense of being coordinate independent. Namely we take all possible global coordinates on $\mathbb{R}^4$ which are standard at infinity, and consider the family of fields defined in each coordinate system. We show that these fields are related by a scalar transformation law. Hence the fields plus transformation law (the covariant field) is naturally arrived at. We also show the scattering operator is independent of coordinates.

Our treatment of these matters seems to shed some light on the general question of how to combine the principles of quantum mechanics with the principle of general covariance (the principle that the dynamics be naturally specified). The importance of effecting this synthesis has been particularly emphasized by Dyson.[16]

II. PRELIMINARIES

A. The metric

We consider Lorentz metrics g on $\mathbb{R}^4$. In standard coordinates the metric is a matrix valued function $g_{\mu\nu}$ on $\mathbb{R}^4$, such that for each $x \in \mathbb{R}^4$ $[x = (x^0,x^1,x^2,x^3) = (x^0,\mathbf{x})]$ the matrix $g_{\mu\nu}(x)$ is symmetric, nondegenerate and has signature (1,3). For each x the metric defines a quadratic form on tangent vectors $v \in \mathbb{R}^4$ $[v = (v^0,v^1,v^2,v^3) = (v^0,\mathbf{v})]$ by $g_x(v,v) = g_{\mu\nu}(x)v^\mu v^\nu$ (summation convention). A tangent vector is called timelike, null, or spacelike according to whether $g_x(v,v)$ is positive, zero, or negative. The nonspacelike vectors form a cone $C_x = \{v: g_x(v,v) \geqslant 0\}$. A smooth curve φ: $[0,1] \to \mathbb{R}^4$ (with $d\varphi/dt \neq 0$) is called nonspacelike if $d\varphi/dt \in C_{\varphi(t)}$. The metric assigns to each nonspacelike curve an interval of proper time

$$\int_0^1 g_{\varphi(t)}\left(\frac{d\varphi}{dt}, \frac{d\varphi}{dt}\right)^{1/2} dt.$$

The cone C_x is actually a double cone, and we assume that there is a continuous splitting $C_x = C_x^+ \cup C_x^-$ into past and future cones, i.e., a time orientation. Nonspacelike curves are either future-directed or past-directed according to $d\varphi/dt \in C^{\pm}_{\varphi(t)}$. The future $J^+(x)$ of $x \in \mathbb{R}^4$ is the set of all points which can be reached from x by a future directed curve, and similarly we define the past $J^-(x)$.

a)Supported by NSF Grant PHY77-21740.

 0022-2488/79/122549-07$01.00

We are interested in metrics which are globally hyperbolic; see Refs. 17–19. One characterization of this property[18] is that for any $x,y\in\mathbb{R}^4$, the lengths of all smooth nonspacelike curves φ between x and y are bounded, where the length is the usual Riemannian length $\int_0^1 |d\varphi/dt|\,dt$. Such metrics have nice properties like $J^{\pm}(x)$ is closed, and $J^{+}(x)\cap J^{-}(y)$ is compact for any x, y.

A special case of the above is the Minkowski metric η corresponding to flat space–time. The metric is $\eta_{\mu\nu}$ $=\mathrm{diag}(1,-1,-1,-1)$ and the cones $C_x^{\pm}$ are

$$V^{\pm}=\{v:(v^0)^2-|\mathbf{v}|^2\geqslant 0,\ \pm v^0\geqslant 0\}. \tag{2.1}$$

This Minkowski metric is globally hyperbolic.

We will be concerned with the following class of metrics:

Definition: A metric g is *admissible* if

(A1) $g=\eta$ off a compact set (including time orientation),

(A2) g is globally hyperbolic.

Now we consider general coordinate systems on $\mathbb{R}^4$. These will always be global and given by a diffeomorphism κ on $\mathbb{R}^4$. We write $\mathrm{diff}(\mathbb{R}^4)$ or just diff for the group of all diffeomorphisms on $\mathbb{R}^4$. We usually also require that κ is the identity off a compact set. All such κ are denoted $\mathrm{diff}_0(\mathbb{R}^4)$ or just diff_0. This is still a rich class. For example, the flow of any smooth vector field with compact support consists of elements diff_0. Using diff_0 instead of diff amounts to giving special treatment to infinity.

The coordinate independent definition of the metric as a tensor field can be formulated as follows. The metric tensor is a function from coordinates on $\mathbb{R}^4$ as given by $\kappa\in\mathrm{diff}$ (or $\kappa\in\mathrm{diff}_0$) to matrix valued functions $g^{\kappa}_{\mu\nu}$ as above, such that for any pair κ_2, κ_1 we have

$$g^{\kappa_2}_{\mu\nu}=D(\kappa_1\circ\kappa_2^{-1})^{\mu'}_{\mu}D(\kappa_1\circ\kappa_2^{-1})^{\nu'}_{\nu}(g^{\kappa_1}_{\mu'\nu'}\circ\kappa_1\circ\kappa_2^{-1}), \tag{2.2}$$

where $(D\kappa)^{\mu'}_{\mu}$ is the Jacobian matrix for κ. In this view the metric tensor is an indexed family of diffeomorphic metrics. We also write this as

$$g^{\kappa_2}=(\kappa_1\circ\kappa_2^{-1})_*\,g^{\kappa_1},$$

where in general for metric g and diffeomorphism κ, $\kappa_*(g)$ is the metric given by

$$(\kappa_*(g))_{\mu\nu}=(D\kappa)^{\mu'}_{\mu}(D\kappa)^{\nu'}_{\nu}(g_{\mu'\nu'}\circ\kappa).$$

A metric tensor will be said to be admissible if g^{κ} is admissible for all $\kappa\in\mathrm{diff}_0$. (Note that this would be impossible for $\kappa\in\mathrm{diff}$ which is one reason why we take diff_0). It is sufficient that it be admissible for some $\kappa\in\mathrm{diff}_0$.

B. The differential operators, distribution densities

The d'Alembertian is defined by $\Box=\nabla_\mu g^{\mu\nu}\nabla_\nu$ where ∇_μ is the covariant derivative, or by

$$\Box=|g|^{-1/2}\partial_\mu|g|^{1/2}g^{\mu\nu}\partial_\nu.$$

Here $|g|=|\det\{g_{\mu\nu}\}|$, $g^{\mu\nu}$ is the inverse matrix to $g_{\mu\nu}$ and $\partial_\mu=\partial/\partial x^\mu$. This is a strictly hyperbolic differential operator which is said to be globally hyperbolic if g is globally hyperbolic.

Let g', g be diffeomorphic metrics, $g'=\kappa_*(g)$, and let $\Box'$, $\Box$ be the associated operator. Since $\Box$ is built from the covariant derivative we have the well-known identity for smooth functions u on $\mathbb{R}^4$

$$\Box'(u\circ\kappa)=(\Box u)\circ\kappa. \tag{2.3}$$

The d'Alembertian naturally defines an operator on scalars as follows. A scalar is a function from coordinates κ (diff or diff_0) to functions u^{κ} such that

$$u^{\kappa_2}=u^{\kappa_1}\circ\kappa_1\circ\kappa_2^{-1}. \tag{2.4}$$

Let $\Box^{\kappa}$ be the family of operators for the tensor g^{κ}. Since $g^{\kappa_2}=(\kappa_1\circ\kappa_2^{-1})_*\,g^{\kappa_1}$ we have by (2.3),

$$\Box^{\kappa_2}u^{\kappa_2}=(\Box^{\kappa_1}u^{\kappa_1})\circ\kappa_1\circ\kappa_2^{-1}.$$

Thus $v^{\kappa}=\Box^{\kappa}u^{\kappa}$ defines a scalar.

The above remarks also hold for distributions. Namely, define the composition of a distribution u and a diffeomorphism κ by

$$\{u\circ\kappa,f\}=\langle u,|\det D\kappa^{-1}|\,f\circ\kappa^{-1}\rangle, \tag{2.5}$$

define a distribution scalar to be a function from diffeomorphisms κ to distributions u^{κ} satisfying (2.4), and then $\Box^{\kappa}$ defines an operator on distribution scalars as above. [(2.3) extends by continuity to distributions.]

Instead of working with scalars we can work with scalar densities. A scalar α-density ($\alpha\in\mathbb{R}$) is a function from coordinates κ to functions or distribution u^{κ} such that

$$u^{\kappa_2}=u^{\kappa_1}\circ\kappa_1\circ\kappa_2^{-1}\,|\det D(\kappa_1\circ\kappa_2^{-1})|^{\alpha}. \tag{2.6}$$

If w^{κ} is a scalar, then $u^{\kappa}=|g^{\kappa}|^{\alpha/2}w^{\kappa}$ is an α-density [use (2.2)] and moreover, every α-density can be written in this way. Instead of $\Box$ we consider the equivalent operators $\Box(\alpha)$ given by

$$\Box(\alpha)=|g|^{\alpha/2}\Box|g|^{-\alpha/2}.$$

Then $\Box(\alpha)$ naturally defines an operator on α-densities in the sense that if u^{κ} is an α-density, then so is $v^{\kappa}=\Box(\alpha)^{\kappa}u^{\kappa}$.

In the following we find it convenient to work with half-densities ($\alpha=\frac{1}{2}$) as well as scalars ($\alpha=0$) translating back and forth as needed. Thus we consider the operators $\bar{\Box}\equiv\Box(\frac{1}{2})=|g|^{1/4}\Box|g|^{-1/4}$ which can be written

$$\bar{\Box}=|g|^{-1/4}\partial_\mu|g|^{1/4}g^{\mu\nu}|g|^{1/4}\partial_\nu|g|^{-1/4}.$$

Note that $\bar{\Box}$ is self-adjoint with respect to the usual inner product $\langle f_1,f_2\rangle=\int f_1(x)f_2(x)\,dx$ on $\mathbb{R}^4$, which is not true for $\Box$. [$\Box$ is self-adjoint if we replace dx by $|g(x)|^{1/2}\,dx$, but the dependence on g is awkward when there is more than one metric, in our case η as well as g.] Another advantage of half-densities is that there is a natural pairing: If $f_1^{\kappa},f_2^{\kappa}$ are half-densities, then $\langle f_1^{\kappa},f_2^{\kappa}\rangle$ is independent of κ, or if u^{κ} is a distribution half-density and f^{κ} is a test function half-density, then $\langle u^{\kappa},f^{\kappa}\rangle$ is independent of κ.

C. The fundamental solutions

Let $\mathscr{D}=\mathscr{D}(\mathbb{R}^4)=C_0^{\infty}(\mathbb{R}^4)$ and $\mathscr{E}=\mathscr{E}(\mathbb{R}^4)$ $=C^{\infty}(\mathbb{R}^4)$ with the usual topologies and with dual spaces $\mathscr{D}'$ (the distributions) and $\mathscr{E}'$ (the distributions with compact support). An operator $T:\mathscr{D}\to\mathscr{D}'$ has a kernel in $\mathscr{D}'(\mathbb{R}^8)$ which we denote $T(x,y)$. (Formally, $\langle f_1,Tf_2\rangle$ $=\int f_1(x)T(x,y)f_2(y)\,dx\,dy$.)

The main result we need is

Theorem 2.1: Let g be an admissible metric. Then there exist operators $E^{\pm}$ mapping $\mathscr{D}$ continuously to $\mathscr{E}$ satisfying:

(a) $E^{\pm}$ are fundamental solution for $(\Box + m^2)$, that is on $\mathscr{D}$,

$$(\Box + m^2)E^{\pm} = E^{\pm}(\Box + m^2) = I.$$

(b) $\text{supp}E^{\pm}(x,y) \subset \{(x,y): x \in J^{\pm}(y)\}$.

Similarly $(\bar{\Box} + m^2)$ has fundamental solutions $\bar{E}^{\pm}$ with this support property.

This follows from a general theorem of Leray[17] for globally hyperbolic differential operators; see also Choquet-Bruhat.[18] The proof is by patching together local fundamental solutions.

Note that although $J^{+}(y)$ can be rather convoluted it cannot extend into the distant past where the metric is Minkowskian. Thus for $f \in \mathscr{D}$, $E^{+}f$ vanishes in the distant past. (A property holds "in the distant past" if it holds on a set $(-\infty, T] \times \mathbb{R}^3$ for some T.)

The following corollaries are not difficult (cf. Refs. 18 and 20). We state them only for $(\bar{\Box} + m^2)$.

Corollary 2.2 (General uniqueness theorem): If $u \in \mathscr{D}'$ satisfies $(\bar{\Box} + m^2)u = 0$ and vanishes in the distant past or future, then $u = 0$.

Corollary 2.3: $\bar{E}^{\pm}$ are the unique operators from $\mathscr{D}$ to $\mathscr{D}'$ which are fundamental solutions and are such that for $f \in \mathscr{D}$, $E^{\pm}f$ vanishes in the distant past (future).

Corollary 2.4: $\bar{E}^{\pm}$ has a continuous extension to an operator from $\mathscr{E}'$ to $\mathscr{D}'$ which is given by $\bar{E}^{\pm} = (\bar{E}^{\mp})'$.

For the Minkowski metric $\bar{\Box} + m^2$ or $\Box + m^2$ reduces to the Klein–Gordon operator $\Box_0 + m^2$ where $\Box_0 = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the wave operator. The fundamental solutions in this case are denoted $E_0^{\pm}$.

We write $\bar{\Box}$ as a (singular) perturbation of $\Box_0$ by $\bar{\Box} = \Box_0 + G$ where

$$G \equiv \bar{\Box} - \Box_0 = (g^{\mu\nu} - \eta^{\mu\nu})\partial_\mu\partial_\nu + \text{lower order}.$$

This second-order differential operator is self-adjoint since $\bar{\Box}, \Box_0$ are self-adjoint. For an admissible metric all the coefficients are in $\mathscr{D}$ and so G is continuous from $\mathscr{E}$ to $\mathscr{D}$ or from $\mathscr{D}'$ to $\mathscr{E}'$.

Proposition 2.5: As operators on $\mathscr{D}$ or $\mathscr{E}'$,

$$\bar{E}^{\pm} = E_0^{\pm} - E_0^{\pm} G\bar{E}^{\pm} = E_0^{\pm} - \bar{E}^{\pm} G E_0^{\pm}.$$

Proof: For $f \in \mathscr{D}$ *let* $v = (\bar{E}^{\pm} - E_0^{\pm} + E_0^{\pm} G\bar{E}^{\pm})f$. A simple calculation shows $(\Box_0 + m^2)v = 0$. Since v vanishes in the distant past or future we have $v = 0$ by Corollary 2.2. This proves the first identity on $\mathscr{D}$ and the second is similar. Taking adjoints gives the identities on $\mathscr{E}'$. Q.E.D.

Finally we consider the transformation properties of $E^{\pm}, \bar{E}^{\pm}$. If $(E^{\pm})'$, $E^{\pm}$ are the fundamental solutions for metrics g', g and $g = \kappa_*(g)$, then by (2.3) and the general uniqueness theorem,

$$(E^{\pm})'(f \circ \kappa) = (E^{\pm}f) \circ \kappa, \quad f \in \mathscr{D}.$$

It follows that the family $(E^{\pm})^{\kappa}$ for a metric tensor g^{κ} defines an operator from $\mathscr{D}$ scalars to $\mathscr{E}$ scalars. Since $\bar{\Box} = |g|^{1/4}\Box|g|^{-1/4}$ we have $\bar{E}^{\pm} = |g|^{1/4}E^{\pm}|g|^{-1/4}$ and so $(\bar{E}^{\pm})_{\kappa}$ defines an operator from $\mathscr{D}$ half-densities to $\mathscr{E}$ half-densities.

III. THE QUANTUM FIELD

We now begin the development of the quantum problem, cf. Wightman.[6,7] The first step is to review the definition of the free field. Let $\mathscr{H} = L_2(\mathbb{R}^3, d\mu)$ be the one particle space where $d\mu(\mathbf{p}) = [2\omega(\mathbf{p})]^{-1}\mathbf{dp}$ and $\omega(\mathbf{p}) = (|\mathbf{p}|^2 + m^2)^{1/2}$. The n-particle space is $\mathscr{H}^n = \otimes_{\text{sym}}^n \mathscr{H}$, and with $\mathscr{H}_0 = \mathbb{C}$ the Fock space is $\mathscr{F} = \oplus_{n=0}^{\infty} \mathscr{H}^n$. Let $\mathscr{F}_0$ be the dense subspace consisting of vectors with a finite number of particles. For $h \in \mathscr{H}$, let $a^*(h), a(h)$ be the usual creation and annihilation operators on $\mathscr{F}_0$ defined so that $[a(h)]^* = a^*(\bar{h})$, so that $a(h)\Omega_0 = 0$ where $\Omega_0 = (1,0,0,\cdots)$ is the no-particle state, and so that the canonical commutation relations (CCR) hold:

$$[a(h_1), a^*(h_2)] = \int h_1(\mathbf{p})h_2(\mathbf{p})\, d\mu(\mathbf{p}).$$

We define $\Pi^{\pm}: \mathscr{S}(\mathbb{R}^4) \to \mathscr{H}$ ($\mathscr{S}$ = Schwartz space) by Fourier transformation (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 then defined by

$$\varphi_0(f) = a(\Pi^{+}f) + a^*(\Pi^{-}f).$$

Then φ_0 satisfies the Klein–Gordon equation $(\Box_0 + m^2)\varphi_0 = 0$ in the sense of distributions, that is,

$$\varphi_0((\Box_0 + m^2)f) = 0$$

and the commmutation relations

$$[\varphi_0(f_1), \varphi_0(f_2)] = \frac{1}{i}\langle f_1, \Delta f_2\rangle,$$

where $\Delta = \Delta_+ + \Delta_-$ is the usual commutator function. We also have $\Delta = E_0 \equiv E_0^{+} - E_0^{-}$.

Now let φ_{in} be a free field as above. We want to define a field operator $\bar{\varphi}$ to satisfy $(\Box_0 + m^2)\bar{\varphi} = 0$ and $\bar{\varphi} = \varphi_{\text{in}}$ in the distant past, and we put $\bar{\varphi} = (1 - \bar{E}^{+}G)\varphi_{\text{in}}$. Since $(1 - \bar{E}^{+}G): \mathscr{D}' \to \mathscr{D}'$ is the adjoint of $(1 - G\bar{E}^{-}): \mathscr{D} \to \mathscr{D}$ what this means precisely is for $f \in \mathscr{D}$,

$$\bar{\varphi}(f) = ((1 - \bar{E}^{+}G)\varphi_{\text{in}})(f) = \varphi_{\text{in}}((1 - G\bar{E}^{-})f).$$

The field $\bar{\varphi}$ is an operator valued distribution, by which we mean that for $f \in \mathscr{D}$, $\bar{\varphi}(f): \mathscr{F}_0 \to \mathscr{F}_0$ and for $\chi, \psi \in \mathscr{F}_0$ the function $f \to (\chi\bar{\varphi}(f)\psi)$ is continuous on $\mathscr{D}$ (these follow from the same properties for φ_{in}).

Theorem 3.1: For an admissible metric:

(a) $\bar{\varphi}$ is the unique operator valued distribution satisfying $(\bar{\Box}_0 + m^2)\bar{\varphi} = 0$ and $\bar{\varphi} = \varphi_{\text{in}}$ in the distant past.

(b) $[\bar{\varphi}(f_1), \bar{\varphi}(f_2)] = (1/i)\langle f_1, \bar{E}f_2\rangle$ where $\bar{E} = \bar{E}^{+} - \bar{E}^{-}$.

Proof: (a) $\bar{\varphi}$ solves the equation because

$$\begin{aligned}\bar{\varphi}((\bar{\Box} + m^2)f) &= \varphi^{\text{in}}((1 - G\bar{E}^{-})(\bar{\Box} + m^2)f) \\ &= \varphi_{\text{in}}((\Box_0 + m^2)f) \\ &= 0.\end{aligned}$$

If f has support in the distant past, then so does $\bar{E}^{-}f = E_0^{-}f$, hence $G\bar{E}^{-}f = 0$ and $\bar{\varphi}(f) = \varphi_{\text{in}}(f)$.

For uniqueness suppose $\bar{\varphi}'$ is another solution and define $u,u' \in \mathscr{D}$ by $\langle u,f\rangle = (\chi,\bar{\varphi}(f)\psi)$ and $\langle u',f\rangle = (\chi, \bar{\varphi}'(f)\psi)$. Then $(\bar{\Box}+m^2)u = 0$, $(\bar{\Box}+m^2)u' = 0$ and both agree with u_{in} the past where $\langle u_{\text{in}},f\rangle = (\chi,\varphi_{\text{in}}(f)\psi)$. By the general uniqueness theorem $u = u'$. Since χ is arbitrary in the dense set $\mathscr{F}_0$ we have $\bar{\varphi}(f)\psi = \bar{\varphi}'(f)\psi$ for $\psi \in \mathscr{F}_0$, i.e., $\bar{\varphi} = \bar{\varphi}'$.

(b) By the free field commutation relations it suffices to show

$$(1 - \bar{E}^{+}G)E_0(1 - G\bar{E}^{-}) = E.$$

However, by Proposition 2.5 we have

$$(1 - \bar{E}^{+}G)E_0^{\pm}(1 - G\bar{E}^{-}) = \bar{E}^{\pm} - \bar{E}^{+}G\bar{E}^{-}.$$

Taking the difference of the plus and minus equations gives the result. Q.E.D.

Now we define $\varphi = |g|^{-1/4}\bar{\varphi}$, that is,

$$\varphi(f) = \bar{\varphi}(|g|^{-1/4}f).$$

Theorem 3.2: For an admissible metric:

(a) φ is the unique operator valued distribution satisfying $(\Box + m^2)\varphi = 0$ and $\varphi = \varphi_{\text{in}}$ in the distant past.

(b) $[\varphi(|g|^{1/2}f_1),\varphi(|g|^{1/2}f_2)] = \langle f_1,|g|^{1/2}Ef_2\rangle$ where $E = E^{+} - E^{-}$.

Proof: Since $\Box = |g|^{-1/4}\bar{\Box}|g|^{1/4}$, part (a) is a direct translation of Theorem 3.1(a). For part (b) the commutator is calculated as $\langle f_1,|g|^{1/4}\bar{E}|g|^{1/4}f_2\rangle$. But since $\bar{E} = |g|^{1/4}E|g|^{-1/4}$ we get the result.

Remarks: (1) $\operatorname{supp}\bar{E}(x,y)$ or $\operatorname{supp}E(x,y)$ are contained in $\{(x,y): x \in J^{+}(y) \cup J^{-}(y)\}$. Thus if $\operatorname{supp}f_1$, $\operatorname{supp}f_2$ are not causally related in the sense that no pair of points can be joined by a nonspacelike curve, then

$$[\bar{\varphi}(f_1),\bar{\varphi}(f_2)] = [\varphi(f_1),\varphi(f_2)] = 0.$$

(2) The form of the commutator in Theorem 3.2(b) was anticipated by Lichnerowicz.[20]

Theorem 3.3: Let g,g' be admissible metrics such that $g' = \kappa_*(g)$ for some $\kappa \in \text{diff}_0$.

(a) If φ,φ' are the fields given by Theorem 3.2, then

$$\varphi' = \varphi \circ \kappa.$$

(b) If $\bar{\varphi},\bar{\varphi}'$ are the fields given by Theorem 3.1, then

$$\bar{\varphi}' = (\bar{\varphi}\circ\kappa)\,|\det D\kappa|^{1/2}.$$

Note: composition in the sense of distributions.

Proof: To prove (a) define $u,u',u_{\text{in}} \in \mathscr{D}'$ by $\langle u,f\rangle = (x,\varphi(f)\psi)$, etc. Then $(\Box + m^2)u = 0$ so $(\Box' + m^2)(u\circ\kappa) = [(\Box + m^2)u]\circ\kappa = 0$. We also have $u\circ\kappa = u_{\text{in}}$ in the distant past. By Corollary 2.2 $u\circ\kappa = u'$ and hence $\varphi\circ\kappa = \varphi'$. Now (b) follows from $\bar{\varphi} = |g|^{1/4}\varphi$, $\bar{\varphi}' = |g'|^{1/4}\varphi'$.

Conclusions: (1) This theorem enables us to define the field as an operator valued distribution scalar or half-density. Namely, let g^{κ} be an admissible metric tensor indexed by $\kappa \in \text{diff}_0$. If we define φ^{κ} to solve $(\Box^{\kappa} + m^2)\varphi^{\kappa} = 0$ as in Theorem 3.2, then Theorem 3.3(a) says that for any $\kappa_2,\kappa_1 \in \text{diff}_0$,

$$\varphi^{\kappa_2} = \varphi^{\kappa_1\circ\kappa_1\circ\kappa_2^{-1}}.$$

If we define $\bar{\varphi}^{\kappa}$ to solve $(\bar{\Box}^{\kappa} + m^2)\bar{\varphi}^{\kappa} = 0$ as in Theorem 3.1, then Theorem 3.3(b) says

$$\bar{\varphi}^{\kappa_2} = (\bar{\varphi}^{\kappa_1\circ\kappa_1\circ\kappa_2^{-1}})\,|\det D(\kappa_1\circ\kappa_2^{-1})|^{1/2}.$$

In either case the dynamics as given by the scalar φ^{κ} or the half-density $\bar{\varphi}^{\kappa}$ has been specified without special choice of coordinates.

(2) We can cast things in a manifestly coordinate free form as follows. Let Φ be the half-density $\bar{\varphi}^{\kappa}$ and let $\mathbb{D}$ be the operator on half-densities defined by the family $\bar{\Box}^{\kappa}$. Then the equations $(\bar{\Box}^{\kappa} + m^2)\bar{\varphi}^{\kappa} = 0$ can be written as

$$(\mathbb{D} + m^2)\Phi = 0.$$

The field Φ can also be regarded as a function from $\mathscr{D}$-half-densities f^{κ} denoted F to operators on $\mathscr{F}_0$ by $\Phi(F) = \bar{\varphi}^{\kappa}(f^{\kappa})$ which is independent of κ, just as $\mathscr{D}$-half-densities themselves have the natural inner product $\langle F_1,F_2\rangle = \langle f_1^{\kappa},f_2^{\kappa}\rangle$. Now if $\mathbb{E}^{\pm},\mathbb{E}$ are the operators on half-densities defined by $(E^{\pm})^{\kappa}$, E^{κ}, we have that the commutator can be written

$$[\Phi(F_1),\Phi(F_2)] = (1/i)\langle F_1,\mathbb{E}F_2\rangle.$$

A similar treatment can be given in the scalar case.

(3) Our treatment is not completely coordinate-free since the data φ_{in} at $x_0 = -\infty$ is given in special Minkowski coordinates, rather than somehow naturally as a scalar or half-density. This is not necessarily a defect. There is no overriding physical principle that the data (like the dynamics) be specified naturally. Moreover, in actual quantum experiments the initial state is prepared (observed) in specific coordinates.

In this connection we remark that it would not seem to be particularly appropriate to try to extend the collection of coordinates from diff_0 to diff: Special treatment has already infected infinity. (Although it would be appropriate to include diffeomorphisms which reduce to a Poincaré transformation off a compact set—this should not be difficult.) We also note that Fadeev[21] advocates a similar restriction on coordinates for quantized gravity.

IV. SINGULARITIES OF $E^{\pm}$

A. Regular metrics

In preparation for the scattering theory in Sec. V we introduce some refinements on the material in Sec. II. The first step is to introduce a new class of metrics.

Definition: A metric g is *regular* if

(R1) $g_{\mu\nu} - \eta_{\mu\nu} \in C_0^{\infty}(\mathbb{R}^4)$,

(R2) $v^0 > 0$ for all $v \neq 0$ in C_x^{+}, all x,

(R3) $(1,0) \in C_x^{\circ +}$ (the interior of C_x^{+}) for all x [i.e., $g_{00}(x) > 0$].

Condition (R1) is the same as (A1). Condition (R2) says the hypersurfaces $x_0 = s = \text{const}$ are spacelike [since the tangent vectors $(0,\mathbf{v})$ are spacelike]. Condition (R3) says the cones do not tip too much, and is only needed for Corollary 4.4 to follow:

Proposition 4.1: A regular metric is admissible.

Proof: We show that (R1), (R2) imply global hyperbolicity (A2). Let $\delta = \inf v^0$ where the infimum is taken over all v with $|v| = 1$, $v \in C_x^{+}$, $x \in \mathbb{R}^4$. Since v^0 is continuous, positive, and $v^0 \geqslant 1/\sqrt{2}$ off a compact set we have $\delta > 0$. Thus for any $v \neq 0$ in C_x^{+} we have $v^0 \geqslant \delta|v| \geqslant \delta|\mathbf{v}|$. Thus all the cones C_x^{+} are contained in the fixed cone $\{v: v^0 \geqslant \delta|\mathbf{v}|\}$.

Now let φ be an arbitrary future-directed curve from x to y ($y_0 > x_0$). Then $|d\varphi/dt| < \delta^{-1}(d\varphi^0/dt)$ and so

$$\int_0^1 \left|\frac{d\varphi}{dt}\right| dt \leqslant (1+\delta^{-1})\int_0^1 \frac{d\varphi^0}{dt}\,dt$$
$$= (1+\delta^{-1})(y_0 - x_0).$$

Thus all such curves have a bounded length and the metric is globally hyperbolic Q.E.D.

Remark: For the scattering theory we consider metrics which are regular for some choice of coordinates $\kappa \in \mathrm{diff}_0$. Such metrics are admissible since regular metrics are admissible. It is not clear whether we are being more restrictive or not. Possibly every admissible metric is regular for some choice of coordinates.. A construction like that of Geroch[19,22] might be a means to prove this.

B. Bicharacteristics

We introduce the cones $\Gamma_x^{\pm}$ dual to $C_x^{\pm}$. with the notation $\xi = (\xi_0,\xi_1,\xi_2,\xi_3) = (\xi_0,\boldsymbol{\xi})$ for $\xi \in \mathbb{R}^4$ we have

$$\Gamma_x^{\pm} = \{\xi : \xi_\mu v^\mu \geqslant 0 \text{ for all } v \in C_x^{\pm}\}.$$

The double cones $\Gamma_x = \Gamma_x^+ \cup \Gamma_x^-$ can also be characterized by

$$\Gamma_x = \{\xi : h_x(\xi,\xi) \geqslant 0\},$$

where

$$h_x(\xi,\xi) = g^{\mu\nu}\xi_\mu\xi_\nu.$$

The conditions for a regular metric can be translated to dual conditions:

(R1′) $g^{\mu\nu} - \eta^{\mu\nu} \in C_0^\infty(\mathbb{R}^4)$,

(R2′) $(1,0) \in \overset{\circ}{\Gamma}{}_x^+$ for all x [i.e., $g^{00}(x) > 0$],

(R3′) $\xi_0 > 0$ for all $\xi \neq 0$ in Γ_x^+, all x.

The dual cones are of interest because $h_x(\xi,\xi)$ is the characteristic polynomial of $(\Box + m^2)$. The bicharacteristics for $(\Box + m^2)$ are the solutions of

$$\dot{x}^\mu = \frac{\partial}{\partial \xi_\mu} h_x(\xi,\xi),$$

$$\dot{\xi}_\mu = -\frac{\partial}{\partial x^\mu} h_x(\xi,\xi),$$

which stay on the manifold

$$\{(x,\xi) : h_x(\xi,\xi) = 0,\ \xi \neq 0\}.$$

This manifold has two components defined by $\xi \in \Gamma_x^{\pm}$. A bicharacteristic is either future or past directed depending on which component it is in.

The bicharacteristic curves are the projections of a bicharacteristic $(x^\mu(t),\xi_\mu(t))$ to $x^\mu(t)$. It is well known that they coincide with the null geodesics for the metric. The latter are solutions of

$$\dot{x}^\mu = v^\mu,$$

$$\dot{v}^\mu = -\Gamma^\mu_{\alpha\beta}(x)v^\alpha v^\beta$$

($\Gamma^\mu_{\alpha\beta}$ = Christoffel symbols), which stay on the manifold

$$\{(x,v) : g_x(v,v) = 0,\ v \neq 0\}.$$

Proposition 4.2: for a regular metric, the bicharacteristics of $\Box + m^2$ satisfy the Duistermaat–Hörmander conditions[9]

(a) No complete bicharacteristic stays over a compact set.

(b) For every compact K, there is a compact K' such that any bicharacteristic curve with endpoints in K is contained in K'.

Proof: (a) We show that no complete null geodesic stays over a compact set. Replace the system above by $\dot{x}^\mu = |v|^{-1}v^\mu$ and $\dot{v}^\mu = -|v|^{-1}\Gamma^\mu_{\alpha\beta}(x)v^\alpha v^\beta$. This changes the speed but not the geometry of the solutions. Since the solutions are nonspacelike curves we have from the proof of Proposition 4.1 that $|v^0| \geqslant \delta|v|$. Thus $|\dot{x}^0| \geqslant \delta$ which gives the result.

(b) A null geodesic with endpoints in K lies in $J^+(K) \cap J^-(K)$. But this is compact by the global hyperbolicity.

C. The wave front set

The wave front set of distribution $u \in \mathscr{D}'(\mathbb{R}^4)$, denoted $WF(u)$, is the complement in $\mathbb{R}^4 \times (\mathbb{R}^4 \setminus 0)$ of the set of all (x^0,ξ_0) such that there exists a neighborhood U of x^0, W of ξ_0 so that for each $f \in C_0^\infty(U)$ and each $N > 0$ there is a constant K so that

$$|(fu)^\sim(\tau\xi)| \leqslant K(1+\tau)^{-N}$$

for all $\xi \in W$, $\tau \geqslant 0$. ("$\sim$" = Fourier transform.) For an operator $T: \mathscr{D}(\mathbb{R}^4) \to \mathscr{D}'(\mathbb{R}^4)$ we define

$$WF'(T) = \{(x,\xi),(y,\eta) : (x,y;\xi,-\eta) \in WF(T(x,y))\}.$$

The next theorem describes the singularities of $E^{\pm}$ in terms of the wave front set.

Theorem 4.3: For a regular metric,

$$WF'(E^{\pm}) \subset \Delta \cup \mathscr{C},$$

where

$$\mathscr{C} = \{\text{all } (x,\xi),(y,\eta) \text{ which lie on the same bicharacteristic}\},$$

$$\Delta = \{(x,\xi),(y,\eta) : x = y,\ \xi = \eta\}.$$

Proof: The proof depends on anther construction of $E^{\pm}$ using Fourier integral operators as developed by Hormander and Duistermaat.[8,9] In fact, our theorem is a special case of a theorem of Duistermaat[10] (Theorem 5.1.6), as we now explain.

Duistermaat's theorem applies to differential operators of arbitrary order on a manifold $\mathbb{R} \times \mathscr{M}$ satisfying the following conditions: (1) The operator is strictly hyperbolic with respect to $\{s\} \times \mathscr{M}$, $s \in \mathbb{R}$ [our condition (R2), (R2′); strict hyperbolicity can be relaxed; see Chazarain[23]]. (2) The operator satisfies the Hörmander–Duistermaat conditions (Proposition 4.2). (3) The domain of dependence between any point x and $\{s\} \times \mathscr{M}$ is compact (e.g., $J^-(x) \cap \{y : y_0 \geqslant 0\}$ is compact; a proof can be had by following the proof of Proposition 4.1).

Under these conditions one constructs solution operators and characterizes their wave front sets. In our case the theorem says there exists a continuous operator $E(s)$ on $\mathscr{C}$ such that $u = E(s)f$ is the unique solution of $(\Box + m^2)u = f$, $u(s,\cdot) = 0$, $(\partial u/\partial x_0)(s,\cdot) = 0$, and the wave front set satisfies

$$WF'(E(s)) \subset \Delta \cup \mathscr{C} \cup \mathscr{C}_s,$$

where $\mathscr{C}_s \subset \{(x,\xi),(y,\eta): y_0 = s\}$. By the uniqueness we have for $f \in \mathscr{D}$ and s sufficiently negative that $E^+ f = E(s)f$. In this situation $\mathscr{C}_s$ has no effect and one can conclude that $WF'(E^+) \subset \Delta \cup \mathscr{C}$. The proof for $WF'(E^-)$ is similar. Q.E.D.

We apply the theorem in the following form:

Corollary 4.4: Let $f_1, f_2 \in \mathscr{D}$. For any N there is a K such that

$$|\langle e^{-ip(\cdot)} f_1, E^{\pm} e^{-iq(\cdot)} f_2 \rangle| \leqslant K(1 + |p|^2 + |q|^2)^{-N}$$

for all $p,q \in V^+$ (or $p,q \in V^-$).

Proof: Let $x,y \in \mathbb{R}^4$ and $p,q \in V^+$, not both zero. We claim that $(x,p;y,-q) \notin WF'(E^{\pm})$. Clearly the point is not on the diagonal Δ so we only need show that $(x,p;y,-q) \notin \mathscr{C}$. Since $-q_0 \leqslant 0$, condition (R3′) says that $(y,-q)$ cannot lie on a future-directed bicharacteristic. Similarly, since $p_0 \geqslant 0$, (x,p) cannot lie on a past-directed bicharacteristic. Thus (x,p) and $(y,-q)$ cannot lie on the same bicharacteristic.

With x, y and p,q as above we now have $(x,y;p,q) \notin WF(E^{\pm}(x,y))$. Therefore there exist neighborhoods U_1, U_2 of x,y, W of (p,q) such that for any $f_1 \in C_0^\infty(U_1)$, $f_2 \in C_0^\infty(U_2)$, and $N > 0$ there is a K so that (in the sense of distributions),

$$\left| \int E^{\pm}(x,y) f_1(x) f_2(y) e^{-i\tau(px+qy)}\, dx\, dy \right| \leqslant K(1+\tau)^{-N}$$

for all $(p,q) \in W$, $\tau \geqslant 0$. By a compactness argument we can extend this bound to hold for any $f_1, f_2 \in \mathscr{D}$ and all $p,q \in V^+$ such that $|p|^2 + |q|^2 = 1$. But this form implies the bound stated in the Corollary.

V. THE SCATTERING OPERATOR

We now are interested in the asymptotic behavior of our field operator $\bar{\varphi}$ (Theorem 3.1) in the distant future. Still assuming an admissible metric we define for $f \in \mathscr{D}$,

$$\varphi_{\text{out}}(f) = ((1 + E_0^- G)\bar{\varphi})(f) = \bar{\varphi}((1 + GE_0^+)f).$$

Then $\bar{\varphi} = \varphi_{\text{out}}$ in the distant future. (And also $\varphi = \varphi_{\text{out}}$ in the distant future). Furthermore it is straightforward to check, as in Theorem 3.1, that φ_{out} satisfies the free field equation and the free field commutation relations. For the moment, however, we cannot conclude that we have a vacuum vector for φ_{out}.

The out field is independent of the choice of coordinates. Namely, suppose we have metrics g, g' such that $g' = \kappa_*(g)$, $\kappa \in \text{diff}_0$. By Theorem 3.3 the associated fields satisfy $\bar{\varphi}' = (\bar{\varphi} \circ \kappa)|\det D\kappa|^{1/2}$. In the distant future this becomes $\varphi'_{\text{out}} = \varphi_{\text{out}}$. Then $\varphi'_{\text{out}} = \varphi_{\text{out}}$ everywhere since as bilinear forms they satisfy the same equation and agree in the future. The result can also be formulated to saying that if $\varphi_{\text{out},\kappa}$ is the out field for a metric tensor g^κ, then $\varphi_{\text{out},\kappa}$ is independent of κ and may be denoted just φ_{out}.

We now ask whether φ_{out} is unitarily equivalent to φ_{in}. Is there a unitary operator S such that $\varphi_{\text{out}} = S^{-1}\varphi_{\text{in}} S$? Clearly, S can be taken to be coordinate independent if it exists at all. To prove existence we may choose special coordinates. We assume that our metric is such that for some choice of coordinates it is regular, i.e., that g^κ is regular for some $\kappa \in \text{diff}_0$. (See the remarks in Sec. 4.1.)

Theorem 5.1: Let φ_{out} be the field for a metric which is regular for some choice of coordinates. Then there exists a (coordinate independent) unitary operator S such that $\varphi_{\text{out}} = S^{-1}\varphi_{\text{in}} S$.

Proof: We assume a regular metric. The out field is given by

$$\varphi_{\text{out}}(f) = (R\varphi_{\text{in}})(f) = \varphi_{\text{in}}(Rf),$$

where $R: \mathscr{D}' \to \mathscr{D}'$ is defined by

$$\begin{aligned} R &= (1 + E_0^- G)(1 - \bar{E}^+ G) \\ &= 1 - E_0(G - G\bar{E}^+ G) \\ &= 1 - E_0 J \end{aligned}$$

and J is defined by

$$J = G - G\bar{E}^+ G.$$

In rewriting R we have replaced E_0^- by $E_0^+ - E_0$ and used Proposition 2.5. Note that $E_0' = -E_0$ and J' are continuous operators on $\mathscr{S}(\mathbb{R}^4)$, hence so is R' and we may continuously extend the definition of φ_{out} to $\mathscr{S}(\mathbb{R}^4)$.

For any h in $\mathscr{S}(\mathbb{R}^3)$, there exists $h^{\pm}$ in $\mathscr{S}(\mathbb{R}^4)$ such that $\Pi^{\pm} h^{\pm} = h$ and $\Pi^{\pm} h^{\mp} = 0$. We define

$$a_{\text{out}}(h) = \varphi_{\text{out}}(h^+), \quad a^*_{\text{out}}(h) = \varphi_{\text{out}}(h^-)$$

Then $[a_{\text{out}}(h)]^* = a^*_{\text{out}}(\bar{h})$ and

$$\varphi_{\text{out}}(f) = a_{\text{out}}(\Pi^+ f) + a^*_{\text{out}}(\Pi^- f).$$

Since φ_{out} satisfies the free field commutation relations it follows that $a_{\text{out}}, a^*_{\text{out}}$ satisfy the CCR. Finally, the relation between the in and out fields becomes

$$a_{\text{out}}(h) = a_{\text{in}}(\Pi^+ R' h^+) + a^*_{\text{in}}(\Pi^- R' h^+).$$

It now suffices to show that there exists a unitary operator satisfying $a_{\text{out}} = S^{-1} a_{\text{in}} S$. According to a standard theorem (e.g., Seiler[24] or Wightman[7]) a CCR preserving transformation as above is unitarily implementable if the off-diagonal operator $h \to \Pi^- R' h^+$ is Hilbert–Schmidt. We now show this.

Since E_0 satisfies

$$\langle f, E_0 h^+ \rangle = (2\pi)^{1/2} i \int \hat{f}(-\omega(\mathbf{q}), -\mathbf{q}) h(\mathbf{q})\, d\mu(\mathbf{q}),$$

and since $R' = 1 + J'E_0$ we have

$$\begin{aligned} &(\Pi^- R' h^+)(\mathbf{p}) \\ &\quad = 2\pi i \int \hat{J}(-\omega(\mathbf{q}), -\mathbf{q}; -\omega(\mathbf{p}), -\mathbf{p}) h(\mathbf{q})\, d\mu(\mathbf{q}), \end{aligned}$$

where $\hat{J}(p,q)$ is the Fourier transform of the kernel $J(x,y)$ of J. Since $J(x,y)$ has compact support, $\hat{J}(p,q)$ is a smooth polynomially bounded function. We show that it is actually rapidly decreasing for $p,q \in V^+$ or $p,q \in V^-$ and hence that $h \to \Pi^- R' h^+$ is Hilbert–Schmidt.

Now G can be written $G = a^{\mu\nu}\partial_\mu \partial_\nu + b^\mu \partial_\mu + c$ where $a^{\mu\nu} = g^{\mu\nu} - \eta^{\mu\nu}$, b^μ and c are in $\mathscr{D}$. Typical terms in $\hat{J}(p,q)$ are thus (up to a constant)

(i) $\hat{a}^{\mu\nu}(p+q) q_\mu q_\nu$,

(ii) $p_\mu p_\nu \langle e^{-ip(\cdot)} a^{\mu\nu}, \bar{E}^+ e^{-iq(\cdot)} a^{\alpha\beta} \rangle q_\alpha q_\beta$,

which come from G and $G\bar{E}^{+}G$ respectively. Now for $p,q\in V^{+}$ we have for any N

$$|\hat{a}^{\mu\nu}(p+q)| \leqslant \mathcal{O}(1+|p+q|^2)^{-N} \leqslant \mathcal{O}(1+|p|^2+|q|^2)^{-N}$$

so that terms of type (i) are rapidly decreasing. Terms of type (ii) are rapidly decreasing by Corollary 4.4 which holds for $\bar{E}^{+}$ as well as E^{+}. [Note: $e^{-ip(\cdot)}$ means the function $x\to\exp[-i(p_0x^0+\mathbf{p}\cdot\mathbf{x})]$ in Corollary 4.4 and $x\to\exp[-i(p_0x^0-\mathbf{p}\cdot\mathbf{x})]$ in (ii), but the conclusion still holds.]

Final remark: If we now define $\Omega_{\text{out}}=S^{-1}\Omega_{\text{in}}$, then $a_{\text{out}}(h)\Omega_{\text{out}}=0$, so we have a vacuum vector for the out field. The states of the form

$$\Phi_{\text{out}}=a^{*}_{\text{out}}(h_1)\cdots a^{*}_{\text{out}}(h_n)\Omega_{\text{out}}$$

may be interpreted as containing n particles with wave functions $h_1,\ldots,h_n$ in the distant future, just as a similar interpretation is held for

$$\Psi_{\text{in}}=a^{*}_{\text{in}}(h'_1)\cdots a^{*}_{\text{in}}(h'_m)\Omega_{\text{in}}$$

in the distant past. Scattering is described by the amplitudes

$$(\Phi_{\text{out}},\Psi_{\text{in}})=(\Phi_{\text{in}},S\Psi_{\text{in}})$$

and features particle creation and annihilation.

Note added in proof: After this manuscript was completed, I learned of the recent lectures of C.J. Isham,[25] which provide a nice mathematical framework for quantum field theory in curved space–time. The present results could be profitably incorporated into that framework.

ACKNOWLEDGMENT

I would like to thank P. Parker for providing me with several useful references.

[1]All the results we obtain will also hold for the equation $(\Box+m^2+\xi R)\varphi=0$ which occurs in the literature.

[2]S. Hawking, "Particle Creation by Black Holes," Commun. Math. Phys. **32**, 199–220 (1975).

[3]B. DeWitt, "Quantum field theory in curved space–time," Phys. Rep. **19C** (1974).

[4]L. Parker, in *Asymptotic Structure of Space–Time*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1977).

[5]Thus we consider asymptotically flat metrics. There are other possibilities such as static metrics which have been studied by C. Moreno, J. Math. Phys. **18**, 2153 (1977); **19**, 92 (1978).

[6]A. Wightman "Relativistic Wave Equations as Singular Hyperbolic Systems," in *Proceedings of the Symposium in Pure Math* (Am. Math. Soc., Providence, R.I., 1973), Vol. 23.

[7]A. Wightman, "Invariant Wave Equations," in *Proceedings of* 1977 *Erice Summer School, Lecture Notes in Physics* (Springer–Verlag, New York, 1978), Vol. 73, pp. 1–101.

[8]L. Hörmander, "Fourier Integral Operators I," Acta Math. **127**, 79–183 (1971).

[9]J. Duistermaat and L. Hörmander,, "Fourier Integral Operators II," Acta Math. **128**, 183–269 (1973).

[10]J. Duistermaat, *Fourier Integral Operators* (Courant Institute Lecture Notes, New York, 1973).

[11]R. Wald, "Existence of the S-matrix in quantum field theory in curved space–time," Chicago preprint.

[12]R. Wald, "Particle Creation by Black Holes," Commun. Math. Phys. **45**, 9–34 (1975).

[13]R. Seiler, "Particles with spin $S\leqslant 1$ in an external field," in *Lecture Notes in Physics* (Springer–Verlag, New York, 1978), Vol. 73.

[14]S.N.M. Ruijsenaars, "Gauge invariance and implementability of the S-operator for spin 0 and spin $\frac{1}{2}$ particles in time-dependent external fields," to appear in J. Func. Anal.

[15]J. Dimock, "Scalar quantum field in an external gauge field," J. Math. Phys. **20**, 1791–96 (1979).

[16]F. Dyson, "Missed Opportunities," Bull. Am. Math. Soc. **78**, 635–52 (1972).

[17]J. Leray, "Hyperbolic differential equations," Princeton Lecture Notes (1953)

[18]Y. Choquet-Bruhat, "Hyperbolic differential equations on a manifold," in *Battelle Rencontres*, edited by DeWitt and Wheeler (Benjamin, New York, 1968); Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, New York, 1977).

[19]S. Hawking and G. Ellis, *The Large Scale Structure of Space–time* (Cambridge Univ., Cambridge, 1973).

[20]A. Lichnerowicz, "Propagateurs, Commutateurs, et Anti-commutateur en Relativité Général," Publication IHES, No. 10 (1961).

[21]L. Fadeev, "Symplectic Structure and Quantization of the Einstein Gravitation Theory," in *Actes due Congrés International des Mathèmaticiens* 1970 (Gauthier–Villars, Paris, 1971), Vol. 3.

[22]R. Geroch, "The domain of dependence," J. Math. Phys. **11**, 437–39 (1970).

[23]J. Chazarain, "Opérateurs hyperbolique à charactéristique de multiplicité constante," Ann. Inst. Fourier **24**, 173–202 (1974).

[24]R. Seiler, Commum. Math. Phys. **25**, 127 (1972).

[25]C.J. Isham, "Quantum Field Theory in Curved Space-Times: A General Mathematical Framework," in *Lecture Notes in Mathematics* (Springer-Verlag, New York, 1978), Vol. 676.

Conformal electrodynamics in Einstein spacetime

W. Heidenreich
Max-Planck Institut zur Erforschung der Lebensbedingungen der wissenschaftlich-technischen Welt, D-813 Starnberg

(Received 3 August 1978; accepted for publication 17 July 1979)

Conformally invariant Maxwell equations in the closed Einstein spacetime are derived, using the diffeomorphism between Einstein spacetime and the universal covering of conformal space. Spatial integration of the charge density yields the total charge 0.

1. INTRODUCTION

Recently, SO(4)-invariant finite energy solutions of the massless φ^4 and the free Yang–Mills theory were found[1], essentially by transforming the nonlinear field equations into static Einstein spacetime, where the SO(4) subgroup of the conformal group SO(4.2) acts as the spatial group of motions.

It is tempting to look along similar lines for solutions of coupled matter and gauge fields, e.g., of the Maxwell–Dirac equations. Here we want to discuss electrodynamics only, using as a guide for finding Maxwell's equations in Einstein spacetime their well-known conformal invariance.

2. ELECTRODYNAMICS IN CONFORMAL SPACE

With conformal space we mean the set of rays

$$x_a \hat{=} \lambda x_a, \quad \lambda > 0, \quad a = 1,...,6, \quad (x_1 ... x_6) \neq (0...0) \tag{1}$$

in a quadric of R^6:

$$x^a x_a = \eta_{ab} x^a x^b = 0;$$
$$\eta_{ab} = (+1, +1, +1, -1, +1, -1). \tag{2}$$

The free electromagnetic field transforms like a conformal vector. The corresponding generators $J = M + S$ are given by

$$J_{ab} = \frac{1}{i}(x_a \partial_b - x_b \partial_a) + S_{ab}\,,$$
$$(S_{ab})^d_c = \frac{1}{i}(\eta_{ac}\eta^d_b - \eta_{bc}\eta^d_a). \tag{3}$$

The second order Casimir operator of SO(4.2) has for the (massless) exceptional degenerate discrete representation[2] the eigenvalue $3(p_0^2 - 1)$, where p_0 is the helicity, in our case $p_0^2 = 1$. Acting on a twice continuously differentiable vector field with degree of homogeneity -1, C_2 has the form

$$(C_2)^d_c a_d = \tfrac{1}{2}(J_{ab}J^{ab})^d_c a_d$$
$$= [\{-(x^a x_a)(\partial^b \partial_b) + (x^a \partial_a)^2 + 4(x_a \partial^a)\}\delta^d_c$$
$$+ \{2(\partial_c x^d - x_c \partial^d) - 2\delta^d_c\} + \{5\delta^d_c\}]a_d = 0. \tag{4}$$

The operators in braces $\{\ \}$ and the operator $x^a\partial_a (x^a \partial_a a_d = -a_d)$ commute with all elements of the Lie algebra (and with one another) and thus are invariant operators which obey an eigenvalue equation in the irreducible representation.

For the first parenthesis we obtain the eigenvalue -3, when restricting the equation to the quadric. Therefore, $\{-x^2\partial^2 + (x\partial)^2 + 4(x\partial)\}a_c = -3a_c$ in general. Now it follows immediately that $(\partial^a \partial_a)a_c = 0$, which must be valid for twice continuously differentiable functions on the quadric.

Equation (4) now gives

$$\partial_c x^d a_d - x_c \partial^d a_d = 0. \tag{4a}$$

Differentiating with respect to ∂^c yields $x^d \partial^c \partial_c a_d - 2\partial^c a_c = 0$ and therefore $\partial^c a_c = 0$. From Eq. (4a) we obtain finally $x^c a_c = c = \text{const}$. So our vector field obeys the following equations:

$$(\partial^a \partial_a)a_b = 0, \quad (x^a \partial_a)a_b = -a_b\,,$$
$$\partial^a a_a = 0, \quad x^a a_a = c = \text{const}. \tag{5}$$

This is Dirac's conformal electrodynamics,[3] if we choose $c = 0$. The last two equations in Eqs. (5) follow from the (MS) term in the Casimir operator.[4]

For the field strengths $f_{ab} = \partial_a a_b - \partial_b a_a$ we obtain

$$x^c \partial_c f_{ab} = -2f_{ab}, \quad \partial^a f_{ab} = 0, \quad x^a f_{ab} = 0. \tag{6}$$

Here the constant c does not appear any longer.

In addition to the usual gauge freedom $a'_b = a_b + \partial_b \Lambda$, we have in Eqs. (5) a conformal gauge freedom

$$a'_a = a_a + x^2 b_a, \quad \text{with} \quad x^a \partial_a b_c = -3b_c,$$
$$\partial^c \partial_c b_a = 0, \quad x^a b_a = 0, \tag{7}$$
$$\partial^a b_a = 0. \tag{7a}$$

It changes the field strengths even on conformal space:

$$f'_{ab} = f_{ab} + 2(x_a b_b - x_b b_a) + x^2(\partial_a b_b - \partial_b b_a). \tag{8}$$

Together with $x^a f_{ab} = 0$ this leaves only six of the 15 f_{ab} independent.

Our explicitly covariant inhomogeneous Maxwell equation on conformal space has the form

$$\partial^a f_{ab} \equiv \partial^a \partial_a a_b - \partial_b \partial^a a_a = j_b. \tag{9}$$

For the current one easily derives

$$x^a \partial_a j_b = -3_{jb}\,, \tag{10}$$
$$x^a j_a = 0, \tag{11}$$
$$\partial^a j_a = 0. \tag{12}$$

The conformal gauge transformation (8) yields

$$\partial^a f'_{ab} = \partial^a f_{ab} - 2x_b \partial^a b_a - 2\partial_b x^a b_a$$
$$+ x^2(\partial^2 b_b - \partial_b \partial^a b_a) = j_b\,, \tag{9a}$$

which stays invariant if b_a fulfills Eqs. (7) and (7a).

Now conformal currents like the spinor current[5] must

 0022-2488/79/122556-03$01.00

have a conformal gauge freedom of the form

$$j'_b = j_b + x_b S \tag{13}$$

on conformal space, where S is a continuous real function with degree of homogeneity -4. (In addition, there are terms of the form $x^2 T_b$, to keep $\partial^b j_b = 0$ invariant. We don't need them for our argument here.)

This current gauge freedom is not in contradiction to the inhomogeneous Maxwell equation on conformal space, if the gauge function b_a obeys

$$\partial^a b_a = -\tfrac{1}{2} S \tag{14}$$

instead of Eq. (7a).

$\partial^a a_a = 0$ in the potential formulation of Eqs. (5) can no longer be valid outside conformal space as it transforms under Eqs. (7) and (14) like

$$\partial^a a'_a = \partial^a a_a + 2x^a b_a + x^2 \partial^a b_a .$$

So our inhomogeneous Maxwell equations on conformal space are Eqs. (9)–(12) and

$$\partial^a a_a = 0, \tag{15}$$

$$x^a a_a = 0, \tag{16}$$

$$x^a \partial_a a_b = -a_b . \tag{17}$$

We choose $x^a a_a = c = 0$ as there is no physical relevance of this constant known to us.

3. ELECTRODYNAMICS IN EINSTEIN SPACE-TIME

Another coordinate system in conformal space is (s_i, α, ρ) with

$$x_i = \frac{\rho}{R} s_i, \quad i = 1,2,3,5; \quad x_4 = \rho \sin\alpha; \quad x_6 = \rho \cos\alpha;$$

$$\rho \mathrel{\hat{=}} \lambda \rho; \quad s^i s_i = R^2; \quad \alpha \in [0, 2\pi). \tag{18}$$

Allowing $\alpha \in (-\infty, +\infty)$, one obtains the universal covering or $S^3 \times R^1$, which is the closed static Einstein spacetime, if we impose the metric

$$ds^2 = -R^2 d\alpha^2 + ds^i ds_i . \tag{19}$$

Equations (18) determine the transformation of the 1-form A:

$$A = a_a dx^a = A_i ds^i + A_\alpha \, d\alpha + A_\rho \, d\rho. \tag{20}$$

Decomposing A_i into a radial part $s_i A^i$ and tangential terms

$$\bar{A}_i = A_i - \frac{s_i}{s^l s_l} s^k A_k, \quad s^i \bar{A}_i = 0, \tag{21}$$

we get

$$a_i = \frac{R}{\rho}\left(\bar{A}_i + \frac{s_i}{s^l s_l} s^k A_k\right),$$

$$a_4 = \frac{1}{\rho}(\sin\alpha s^i A_i - \rho \sin\alpha A_\rho - \cos\alpha A_\alpha),$$

$$a_6 = \frac{1}{\rho}(\cos\alpha s^i A_i - \rho \cos\alpha A_\rho + \sin\alpha A_\alpha). \tag{22}$$

In just the same way the vector ∂_a is transformed into $\bar{\partial}_i$, ∂_α, and ∂_ρ; the transition from the 1-form components j_a to $\bar{J}_i$, J_α and J_ρ has an additional factor $1/\rho^2$ at the right side to make $J_{i,\alpha}(s_i, \alpha)$ independent of ρ.

Now the field equations (9–12) and (15–17) can be given in the new coordinates. The notation becomes especially handy if one uses the field strengths

$$\bar{F}_{ij} = F_{ij} - \frac{1}{s^l s_l} s^k (s_j F_{ik} - s_i F_{jk})$$

$$= \bar{\partial}_i \bar{A}_j - \bar{\partial}_j \bar{A}_i - \frac{1}{s^l s_l}(s_i \bar{A}_j - s_j \bar{A}_i); \tag{23}$$

$$\bar{F}_{i\alpha} = F_{i\alpha} - \frac{s_i}{s^l s_l} s^k F_{k\alpha}, \quad s^j \bar{F}_{ij} = s^j \bar{F}_{i\alpha} = 0.$$

We obtain from Eqs. (16), (11), (15), and (12)

$$A_\rho = J_\rho = 0, \tag{24}$$

$$s^i A_i = \tfrac{1}{2}(\partial_\alpha A_\alpha - R^2 \bar{\partial}^i \bar{A}_i), \tag{25}$$

$$R^2 \bar{\partial}^i \bar{J}_i = \partial_\alpha J_\alpha . \tag{26}$$

If we put $a = i$ in Eq. (9), it follows that

$$\bar{J}_i = R^2 \bar{\partial}^j \bar{F}_{ji} + \partial_\alpha \bar{F}_{i\alpha}, \tag{27}$$

and with $a = 4,6$, we have

$$J_\alpha = R^2 \bar{\partial}^j \bar{F}_{j\alpha} , \tag{28}$$

$$s^i J_i = R^2 \partial^k s^j F_{kj} + \partial_\alpha s^k F_{k\alpha} . \tag{29}$$

$\bar{J}_i, J_\alpha, \bar{F}_{ij}, \bar{F}_{i\alpha}$, and Eq. (25) stay invariant under conformal gauge transformations, but not Eq. (29):

$$s^i J'_i = s^i J_i + s^i s_i \rho^4 S.$$

So $s^i J_i$ is a pure gauge component of the conformal current. It can always be made to vanish with an appropriate conformal gauge transformation. Equation (29) then yields

$$R^2 \partial^k s^j F_{kj} + \partial_\alpha s^k F_{k\alpha} = 0,$$

and if one eliminates all $s^i A_i$ terms with the help of Eq. (25),

$$(R^2 \bar{\partial}_i \bar{\partial}^i - \partial_\alpha^2)(\partial_\alpha A_\alpha - R^2 \bar{\partial}^j \bar{A}_j) - 4\partial_\alpha A_\alpha = 0. \tag{30}$$

This is the conformally invariant gauge condition, which corresponds to $\Box \partial^\nu A_\nu = 0$ in Minkowski spacetime[6]; together with Eqs. (23), (27), and (28) it forms conformal electrodynamics in Einstein spacetime. Equation (26) is the continuity equation for the current.

In the usual generalization of Maxwell's equations into curved spacetime[7], the Lorentz condition gets $A^\nu{}_{;\nu} = 0$, or in the present notation

$$\partial_\alpha A_\alpha - R^2 \bar{\partial}^i \bar{A}_i = 0. \tag{31}$$

with our conformal gauge condition this yields

$$\partial_\alpha A_\alpha = 0, \quad R^2 \bar{\partial}^i \bar{A}_i = 0, \tag{32}$$

which in general cannot be obtained with a gauge transformation. So the covariant generalization of the Lorentz condition cannot be imposed on the conformal electrodynamics in Einstein spacetime.

4. CHARGE OF THE UNIVERSE

If there is one charge in an otherwise empty Minkowski spacetime, one can imagine the electrical field as rays going to infinity. This is impossible in the closed Einstein spacetime; the field lines have to end at an opposite charge; the total charge must be 0. Is the naive picture correct? To de-

cide this question one can integrate Eq. (28) over S^3 directly or use polar coordinates

$$s_1 = r\sin\chi\sin\vartheta\cos\varphi; \quad s_3 = r\sin\chi\cos\vartheta;$$
$$s_2 = r\sin\chi\sin\vartheta\sin\varphi; \quad s_5 = r\cos\chi. \tag{33}$$

They give in an unique way formulas for ∂_χ, $F_{\chi\alpha}$, etc. Equation (28) becomes

$$\sin^2\chi\sin\vartheta J_\alpha$$
$$= \partial_\chi(\sin^2\chi\sin\vartheta F_{\chi\alpha}) + \partial_\vartheta(\sin\vartheta F_{\vartheta\alpha}) = \partial_\varphi\left(\frac{1}{\sin\vartheta}F_{\varphi\alpha}\right), \tag{34}$$

and integrating over space S^3:

$$Q = R^3\int_{S^3}\sin^2\chi\sin\vartheta J_\alpha\, d\chi\wedge d\vartheta\wedge d\varphi$$
$$= R^3\int_{\partial S^3}\Big[(\sin^2\chi\sin\vartheta F_{\chi\alpha})\,d\vartheta\wedge d\varphi$$
$$+ (\sin\vartheta F_{\vartheta\alpha})\,d\varphi\wedge d\chi + \left(\frac{1}{\sin\vartheta}F_{\varphi\alpha}\right)d\chi\wedge d\vartheta\Big] = 0. \tag{35}$$

Here we used Stokes theorem for forms $\int_{S^3}d\varphi = \int_{\partial S^3}\varphi$, and the fact that the boundary of S^3 is ϕ.

One could have obtained this result more easily from the usual covariant generalization of Maxwell's equations:

$$F^{\mu\nu}{}_{;\nu} = J_\mu \quad \text{or} \quad d{*}dA = {*}J. \tag{36}$$

However the conformally invariant spin 0 equation in Minkowski spacetime will yield $\varphi^{;\nu}{}_{;\mu} = 0$ in Einstein spacetime if one follows the covariant generalization procedure[7], and this is not the conformally invariant equation $\varphi^{;\nu}{}_{;\nu} - (1/R^2)\varphi = 0$. A similar discepancy from Eq. (36) would not amount to the total charge 0.

5. CONCLUSIONS

We have shown how, in a conformally invariant gauge theory, global topological properties of the underlying spacetime can fix the global charge connected with an inner symmetry group.

Simultaneously, the original aim to find classical solutions of the coupled Maxwell–Dirac equations cannot be achieved. This is so because the charge density of the Dirac current is definite, and so the "total charge" of a nonvanishing spinor cannot be 0. The same problem, which leads to quantization with anticommutators, doesn't allow (commutating) c-number solutions.

A last remark: Every charge in Einstein spacetime can be assigned to an opposite charge. Thus, it is possible to formulate electrodynamics in a more symmetric way without electric monopoles, using electric and magnetic dipoles only.

This has been pointed out to the author by L. Castell, with whom many discussions are acknowledged.

[1]L. Castell, Phys. Rev. D **6**, 536 (1972); Acta Phys. Austr. Suppl. **4**, 847 (1972); V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. B **65**, 163 (1976).
[2]T. Yao, J. Math. Phys. **9**, 1615 (1968).
[3]P.A.M. Dirac, Ann. Math. **37**, 429 (1936).
[4]W. Heidenreich, Diplom thesis, TU München, 1977.
[5]H.A. Kastrup, Phys. Rev. **150**, 1183 (1966).
[6]D.M. Mayer, J. Math. Phys. **16**, 884 (1975); F. Bayen and M. Flato, J. Math. Phys. **17**, 1112 (1976).
[7]For example, C.M. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1972), p. 569.

Low-temperature thermodynamics of a weakly interacting Bose gas confined to restricted geometries[a)]

C. S. Zasada[b)] and R. K. Pathria
Department of Physics, University of Waterloo, Waterloo, Ontario, Canada

(Received 9 October 1978; accepted for publication 8 June 1979)

The low-temperature internal energy and specific heat of a hard sphere Bose gas confined to a general cuboidal enclosure are calculated under periodic boundary conditions. The special cases of a film and a channel geometry are considered in some detail. It is shown that, in the appropriate limits, one recovers the recent results of Pajkowski and Pathria, and of Zasada and Pathria, for a system of phonons confined to different geometries. Systematic applications of the Poisson summation formula enable us to obtain strongly convergent expansions, in both $\lambda/L \ll 1$ and $\lambda/L \gg 1$ regimes, and to see how the magnitude of the finite size corrections in the asymptotic limit ($L/\lambda \gg 1$) is directly linked with the dimensionality of the system. Special attention is paid to examining the combined influence of the dispersion in the phonon spectrum of the system and the finiteness of the enclosure.

I. INTRODUCTION

Recent interest in the study of finite size effects in various physical systems, notably superfluid helium, has prompted several workers into carrying out theoretical investigations of such effects in a Bose–Einstein gas. Most of these investigations have centered on a gas of noninteracting particles confined to a cuboidal geometry and subject to a variety of boundary conditions.[1–5] Clearly, the relevance of these studies to the problem of actual superfluid helium is bound to be limited unless one includes the influence of interparticle interactions. The first indirect step in this direction was taken by Padmore[6] who suggested that the finite size effects in superfluid helium may be analyzed within the framework of Landau's quasiparticle picture, provided that the smallest dimension of the container is much larger than the healing length, which is of the order of 1–2 Å. In this spirit, he and other authors[7–9] have examined the thermodynamics of a phonon gas in different geometries under various boundary conditions; the results thus obtained should be of direct relevance to liquid helium below 0.5 K. In the same spirit we undertook to investigate the thermodynamics of a hard sphere Bose gas confined to an arbitrary cuboidal enclosure. This investigation is based on the energy spectrum derived by Brueckner and Sawada,[10] and emphasis is laid on examining the combined influence of the dispersion in the phonon spectrum of the system and the finiteness of the enclosure.

II. PRELIMINARIES

The Brueckner–Sawada spectrum for a hard sphere Bose gas is of the form

$$\epsilon(\mathbf{k}) = \frac{\hbar^2 k}{2m}\left(k^2 + \frac{2\sigma}{a^2}\frac{\sin(ka)}{(ka)}\right)^{1/2}, \tag{1}$$

where m is the particle mass and the parameter σ (which is a function of the particle density ρ and the hard sphere diameter a) is given by the implicit relation

$$4\pi^2\rho a^3 = \sigma\int_0^\infty \frac{x\sin^2 x}{x^3 + \sigma\sin x}\,dx. \tag{2}$$

If the mean momentum of the excitations is so small that the term $\sin(ka)/(ka)$ is only slightly different from unity, one deals with a monotonic, dispersive spectrum, viz.,

$$\epsilon(\mathbf{k}) = \frac{\hbar k}{2m}(\hbar^2 k^2 + r^2)^{1/2}, \tag{3}$$

where

$$r^2 = 2\sigma(\hbar/a)^2\,\eta\,; \tag{4}$$

here, η denotes the mean value of the oscillating term $\sin(ka)/(ka)$ averaged over an assembly of the excitations in question. Since we are interested in the low-temperature properties of the system, we can safely assume Eq. (3) to be the operative form of the energy spectrum. In the phonon limit,

$$\eta = \sum_{n=0}^{\infty}(-1)^n(n+1)\frac{\zeta(2n+3)}{\zeta(3)}\left(\frac{a}{\lambda}\right)^{2n} \quad \left(\lambda = \frac{\hbar c}{kT}\right) \tag{5}$$

$$\simeq 1 - \frac{1}{\zeta(3)}\left[1 - \left(1 + \frac{a^2}{\lambda^2}\right)^{-2}\right], \tag{6}$$

which is essentially equal to unity because, for all practical purposes, $(a/\lambda) \ll 1$. The speed of sound c is related to the parameters of the spectrum through the relation

$$c = \lim_{p\to 0}[\epsilon(p)/p] = (\sigma/2)^{1/2}(\hbar/ma)\,\eta^{1/2}\,; \tag{7}$$

however, the temperature dependence of η is so weak that we may neglect it throughout this analysis and use, for c, the empirical value of 238 m/sec [11] for liquid ^{4}He.

We can now write down thermodynamic expressions for the internal energy U and the specific heat C_V of these excitations, viz.,

$$U = \sum_{\mathbf{k}}\epsilon(\mathbf{k})\{\exp[\beta\epsilon(\mathbf{k})] - 1\}^{-1} \tag{8}$$

[a)]Work supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/79/122559-08$01.00

and

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V, \tag{9}$$

where $\epsilon(\mathbf{k})$ is given by (3). The bulk situation has been studied in detail by Pathria and Singh,[12] who obtained the following results:

$$U = \frac{\pi V r^5}{16 m h^3} \sum_{n=1}^{\infty} [S_{5/2}(nx) - S_{3/2}(nx) - 2S_{1/2}(nx)] \tag{10}$$

and

$$C_V = \frac{\pi V r^5}{32 m h^3 T} \sum_{n=1}^{\infty} (nx)[C_{7/2}(nx) - C_{5/2}(nx) - 3C_{3/2}(nx) + 3C_{1/2}(nx)], \tag{11}$$

where $S_\nu(z)$ and $C_\nu(z)$ are the Lommel functions, defined by[13]

$$S_\nu(z) = \int_0^\infty \exp(-z\sinh t)\sinh(\nu t)\,dt = -S_{-\nu}(z) \tag{12}$$

and

$$C_\nu(z) = \int_0^\infty \exp(-z\sinh t)\cosh(\nu t)\,dt = C_{-\nu}(z), \tag{13}$$

while $x = r^2/4mkT$. In the weakly dispersive region ($x \gg 1$), one may employ the asymptotic expansions

$$S_\nu(z) = \nu\left(\frac{1}{z^2} - \frac{(2^2-\nu^2)}{z^4} + \frac{(2^2-\nu^2)(4^2-\nu^2)}{z^6} - \cdots\right), \tag{14}$$

$$C_\nu(z) = \left(\frac{1}{z} - \frac{(1^2-\nu^2)}{z^3} + \frac{(1^2-\nu^2)(3^2-\nu^2)}{z^5} - \cdots\right), \tag{15}$$

and obtain

$$U = V\frac{4\pi^5(kT)^4}{15h^3c^3}\left[1 - \frac{25}{21}\pi^2\left(\frac{kT}{mc^2}\right)^2 + \cdots\right] \tag{16}$$

and

$$c_v = \frac{C_V}{V} = \frac{16\pi^5 k^4 T^3}{15h^3c^3}\left[1 - \frac{25}{14}\pi^2\left(\frac{kT}{mc^2}\right)^2 + \cdots\right]. \tag{17}$$

The leading terms in these expressions represent the bulk situation in a standard phonon gas while the subsequent terms represent the bulk dispersive corrections.[14,15]

III. THE CUBOIDAL GEOMETRY ($L_1 \times L_2 \times L_3$)

For finite systems, we must aim at evaluating the discrete sum-over-states appearing in Eq. (8) in a manner which yields the result as a bulk term plus a set of finite size corrections. The method most suited for this purpose is the application of the Poisson summation formula (PSF) which, in a d-dimensional space, states that

$$\sum_{n_1,\ldots,n_d=-\infty}^{\infty} f(n_1,\ldots,n_d) = \sum_{q_1,\ldots,q_d=-\infty}^{\infty} \mathcal{F}(q_1,\ldots,q_d), \tag{18}$$

where

$$\mathcal{F}(q_1,\ldots,q_d) = \int_{-\infty}^{\infty} f(\mathbf{r})\, e^{-2\pi i \mathbf{q}\cdot\mathbf{r}}\, d^d r. \tag{19}$$

It can readily be seen that the term with $\mathbf{q} = 0$ on the right-hand side of (18) is precisely the result one would obtain by replacing the original summation over $\mathbf{n}$ by an integration. It then follows that the term $\mathcal{F}(0)$ would correspond to the bulk situation while terms $\mathcal{F}(\mathbf{q})$ with $\mathbf{q} \neq 0$ would give the finite size corrections. An added benefit of the PSF is that, almost invariably, it converts a slowly convergent sum into a rapidly converging one. For instance, Jacobi's imaginary transformation for theta functions is just an application of the one-dimensional PSF to the summand $f(n) = \exp(-n^2 t)$, viz.,

$$\sum_{n=-\infty}^{\infty} e^{-n^2 t} = (\pi/t)^{1/2} \sum_{q=-\infty}^{\infty} e^{-\pi^2 q^2/t}.$$

For small values of t, the left-hand side of this identity converges quite slowly while the right-hand side converges very rapidly. For further examples, see Chaba and Pathria.[16]

Returning to the problem at hand, if we subject our system to periodic boundary conditions, we have

$$k_i = (2\pi/L_i)\, n_i, \quad n_i = 0, \pm 1, \pm 2, \cdots \quad (i = 1,2,3).$$

Accordingly, Eq. (8) becomes

$$\beta U = \pi\Lambda^2 \sum_{j=1}^{\infty} \sum_{n_{1,2,3}=-\infty}^{\infty} n'(n'^2+\gamma^2)^{1/2} \exp[-\pi j \Lambda^2 n'(n'^2+\gamma^2)^{1/2}], \tag{20}$$

where $\Lambda = h/(2\pi mkT)^{1/2}$, $\gamma = r/h = 2mc/h$ and $\mathbf{n}' = (n_1/L_1, n_2/L_2, n_3/L_3)$. Application of the three-dimensional PSF to this expression gives

$$\beta U/V = 2\pi\Lambda^2 \sum_{j=1}^{\infty} \sum_{q_{1,2,3}=-\infty}^{\infty} \frac{1}{q'} \int_0^\infty \xi^2(\xi^2+\gamma^2)^{1/2} \exp[-\pi\Lambda^2 j\xi(\xi^2+\gamma^2)^{1/2}] \sin(2\pi q'\xi)\, d\xi,$$

where $\mathbf{q}' = (q_1 L_1, q_2 L_2, q_3 L_3)$. Substituting $\xi = \gamma \sinh(\frac{1}{2}t)$, this becomes

$$\beta u = \frac{\pi}{4} \Lambda^2 \gamma^4 \sum_{j=1}^{\infty} \sum_{q_{1,2,3}=-\infty}^{\infty} \frac{1}{q'} \int_0^{\infty} \sinh^2 t \exp(-jx \sinh t) \sin|2\pi\gamma q' \sinh(\tfrac{1}{2}t)| \, dt,$$

where $u = U/V$ and $x = (\pi/2)\Lambda^2\gamma^2 = r^2/4mkT = mc^2/kT$. This integral is not tabulated but, through the use of the generating function for Bessel functions, we can establish that

$$\sin[2\pi\gamma q' \sinh(\tfrac{1}{2}t)] = 2 \sum_{k=0}^{\infty} (-1)^k I_{2k+1}(2\pi\gamma q') \sinh(k+\tfrac{1}{2})t, \tag{21}$$

where $I_\nu(z)$ are the modified Bessel functions. Employing this expansion along with Eq. (12), we get

$$\beta u = \frac{\pi}{8} \Lambda^2 \gamma^4 \sum_{q_{1,2,3}=-\infty}^{\infty} \frac{1}{q'} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} (-1)^k I_{2k+1}(2\pi\gamma q') \left[S_{k+5/2}(jx) + S_{k-3/2}(jx) - 2S_{k+1/2}(jx)\right]. \tag{22}$$

As expected, the $(\mathbf{q} = 0)$ term here is precisely the bulk result, Eq. (10), of Pathria and Singh.

While Eq. (22) is exact for a hard sphere Bose gas with energy spectrum (3) its utility is limited because of the several summations yet to be done. However, if we expand the Lommel functions appropriately, the j and k summations can actually be carried out. We will do this only for the leading dispersive corrections to the phonon results although, in principle, this can be carried through to any order in the dispersion parameter $1/x$. For $x \gg 1$, we have

$$C_{\nu-1}(x) - C_{\nu+1}(x) = 2\frac{\partial}{\partial x} S_\nu(x) = -4\nu\left(\frac{1}{x^3} - \frac{2(2^2-\nu^2)}{x^5} + \frac{3(2^2-\nu^2)(4^2-\nu^2)}{x^7} - \cdots\right) \tag{23}$$

and

$$\begin{aligned} S_{\nu+2}(x) + S_{\nu-2}(x) - 2S_\nu(x) &= 2\frac{\partial}{\partial x}\left[C_{\nu-1}(x) - C_{\nu+1}(x)\right] \\ &= 8\nu\left(\frac{1\cdot3}{x^4} - \frac{2\cdot5\cdot(2^2-\nu^2)}{x^6} + \frac{3\cdot7\cdot(2^2-\nu^2)(4^2-\nu^2)}{x^8} - \cdots\right). \end{aligned} \tag{24}$$

Using (24), the j-summation in (21) can be readily carried out:

$$\sum_{j=1}^{\infty} \left\{S_{\nu+2}(jx) + S_{\nu-2}(jx) - 2S_\nu(jx)\right\} = 8\nu\left(\frac{1\cdot3}{x^4}\zeta(4) - \frac{2\cdot5\cdot(2^2-\nu^2)}{x^6}\zeta(6) + \frac{3\cdot7\cdot(2^2-\nu^2)(4^2-\nu^2)}{x^8}\zeta(8) - \cdots\right); \tag{25}$$

in our problem $\nu = k + \frac{1}{2}$. For carrying out the k-summations, we have to consider the sums

$$\Sigma_p(z) = \sum_{k=0}^{\infty} (-1)^k (2k+1)^{2p+1} I_{2k+1}(z) \quad (p = 0,1,2,\cdots). \tag{26}$$

These can be done by using successive differentiations of the identity (21). The final result can be summarized as

$$\Sigma_p(z) = \sum_{l=0}^{p} (-1)^l \frac{1}{(2l+1)!} \mathscr{L}_{2l+1}^{(2p+1)} \left(\frac{z}{2}\right)^{2l+1},$$

where

$$\mathscr{L}_{2l+1}^{(2p+1)} = \sum_{m=0}^{l} (-1)^m \binom{2l+1}{m} (2l+1-2m)^{2p+1}.$$

Some of the coefficients $\mathscr{L}_{2l+1}^{(2p+1)}$ have been tabulated in Table I; they are sufficient for evaluating the first six Σ-polynomials. We get

$$\Sigma_0(z) = \tfrac{1}{2}z,$$

TABLE I. Values of $\mathscr{L}_{2l+1}^{(2p+1)}/(2l+1)!$

p \ l	0	1	2	3	4	5
0	1	—	—	—	—	—
1	1	4	—	—	—	—
2	1	40	16	—	—	—
3	1	364	560	64	—	—
4	1	3,280	15,456	5,376	256	—
5	1	29,524	399,520	325,248	42,240	1,024

$$\Sigma_1(z) = \tfrac{1}{2}(z - z^3),$$
$$\Sigma_2(z) = \tfrac{1}{2}(z - 10z^3 + z^5),$$
$$\Sigma_3(z) = \tfrac{1}{2}(z - 91z^3 + 35z^5 - z^7),$$
$$\Sigma_4(z) = \tfrac{1}{2}(z - 820z^3 + 966z^5 - 84z^7 + z^9),$$
$$\Sigma_5(z) = \tfrac{1}{2}(z - 7{,}381z^3 + 24{,}970z^5 - 5{,}082z^7 + 165z^9 - z^{11}).$$

These polynomials are sufficient for determining the finite size effects in a system of phonons as well as the leading contributions to finite size effects attributable to dispersion.

Substituting the foregoing results into Eq. (22), we finally obtain

$$\beta u = (\beta u)_{\text{bulk}} + \frac{\pi^2}{2}\Lambda^2\gamma\left(\frac{\gamma}{x}\right)^4\left\{\sum_{q_{1,2,3}=-\infty}^{\infty}{}' \left[1\cdot3\cdot\zeta(4) - 2\cdot5\cdot\zeta(6)\left(\frac{\pi\gamma}{x}q'\right)^2 + 3\cdot7\cdot\zeta(8)\left(\frac{\pi\gamma}{x}q'\right)^4 - \cdots\right]\right.$$
$$\left. - \frac{1}{16x^2}\sum_{q_{1,2,3}=-\infty}^{\infty}{}' \left[2\cdot3\cdot4\cdot5^2\zeta(6) - 4\cdot5\cdot6\cdot7^2\cdot\zeta(8)\left(\frac{\pi\gamma}{x}q'\right)^2 + 6\cdot7\cdot8\cdot9^2\cdot\zeta(10)\left(\frac{\pi\gamma}{x}q'\right)^4 - \cdots\right] + \cdots\right\}. \tag{27}$$

The leading set of correction terms here is of the form

$$(\pi^2/2)\Lambda^2\gamma(\gamma/x)^4\sum_{q_{1,2,3}=-\infty}^{\infty}{}' J((\pi\gamma/x)q'), \tag{28}$$

where

$$J(a) = \frac{1}{2}\sum_{n=0}^{\infty}(-1)^n(2n+2)(2n+3)\zeta(2n+4)a^{2n} = \frac{1}{2a}\int_0^\infty \frac{y^2}{(e^y-1)}\left(\sum_{n=0}^{\infty}(-1)^n\frac{(ay)^{2n+1}}{(2n+1)!}\right)dy$$
$$= \frac{1}{2a}\int_0^\infty \frac{y^2\sin(ay)}{(e^y-1)}dy = \frac{1}{2a}\left(\frac{1}{a^3} - \pi^3\coth(\pi a)\operatorname{csch}^2(\pi a)\right). \tag{29}$$

Similarly, the second set of terms is of the form

$$-(\pi^2/32)\Lambda^2(\gamma/x^2)(\gamma/x)^4\sum_{q_{1,2,3}=-\infty}^{\infty}{}' K((\pi\gamma/x)q'), \tag{30}$$

where

$$K(a) = \sum_{n=0}^{\infty}(-1)^n(2n+2)(2n+3)(2n+4)(2n+5)^2\zeta(2n+6)a^{2n} = \int_0^\infty \frac{y^5\cos(ay)}{(e^y-1)}dy + \frac{4}{a}\int_0^\infty \frac{y^4\sin(ay)}{(e^y-1)}dy$$
$$= \frac{12}{a^6} + \frac{16\pi^5}{a}(2\coth\pi a\operatorname{csch}^4\pi a + \coth^3\pi a\operatorname{csch}^2\pi a)$$
$$- 4\pi^6(11\coth^2\pi a\operatorname{csch}^4\pi a + 2\coth^4\pi a\operatorname{csch}^2\pi a + 2\operatorname{csch}^6\pi a). \tag{31}$$

Now, in terms of the mean thermal wavelength of the phonons λ ($= \hbar c/kT$), we have $\Lambda^2 = 2\pi\lambda^2/x$ and $\gamma = x/\pi\lambda$, where $x = mc^2/kT$. Employing these relations in the foregoing equations, we finally obtain

$$\beta u = (\beta u)_{\text{bulk}} + \frac{1}{2\pi^2\lambda^3}\sum_{q_{1,2,3}=-\infty}^{\infty}{}' \left\{\left(\frac{\lambda}{q'}\right)^4 - \pi^3\left(\frac{\lambda}{q'}\right)\coth\left(\frac{\pi}{\lambda}q'\right)\operatorname{csch}^2\left(\frac{\pi}{\lambda}q'\right)\right\} - \frac{1}{4\pi^2\lambda^3}\left(\frac{kT}{mc^2}\right)^2$$
$$\times\sum_{q_{1,2,3}=-\infty}^{\infty}{}' \left\{3\left(\frac{\lambda}{q'}\right)^6 + 4\pi^5\left(\frac{\lambda}{q'}\right)\left[2\coth\left(\frac{\pi}{\lambda}q'\right)\operatorname{csch}^4\left(\frac{\pi}{\lambda}q'\right)\right] + \coth^3\left(\frac{\pi}{\lambda}q'\right)\operatorname{csch}^2\left(\frac{\pi}{\lambda}q'\right)\right.$$
$$\left. - \pi^6\left[11\coth^2\left(\frac{\pi}{\lambda}q'\right)\operatorname{csch}^4\left(\frac{\pi}{\lambda}q'\right) + 2\coth^4\left(\frac{\pi}{\lambda}q'\right)\operatorname{csch}^2\left(\frac{\pi}{\lambda}q'\right) + 2\operatorname{csch}^6\left(\frac{\pi}{\lambda}q'\right)\right]\right\} + \cdots. \tag{32}$$

Here, $(\beta u)_{\text{bulk}}$ is precisely the bulk result given by Eq. (10), the second term is precisely the finite size correction obtained from an *ab initio* calculation for a standard, nondispersive phonon gas ($\epsilon = pc$) in a cuboidal enclosure under periodic boundary conditions,[9] while the subsequent terms represent finite size corrections arising from the dispersive nature of the energy spectrum (3). Since the finite size corrections in the standard phonon gas have been discussed at length by Pajkowski and Pathria[8] and by ourselves,[9] we shall concentrate here mainly on the dispersive effects.

IV. DISPERSIVE EFFECTS IN A FILM GEOMETRY ($\infty \times \infty \times L$)

In the limit of a film geometry, the summations appearing in (32) become one-dimensional, so that

$$(\beta u)_{\text{disp}} = -\frac{1}{\lambda^3}\left(\frac{kT}{mc^2}\right)^2\left(\frac{3}{2\pi^2}\zeta(6)\left(\frac{\lambda}{L}\right)^6 + \sum_{q=1}^{\infty}\left\{2\pi^3\left(\frac{\lambda}{L}\right)\right.\right.$$
$$\left.\left.\times\left(\frac{2\coth((\pi L/\lambda)q)\operatorname{csch}^4((\pi L/\lambda)q) + \coth^3((\pi L/\lambda)q)\operatorname{csch}^2((\pi L/\lambda)q)}{q}\right) - \frac{\pi^4}{2}\left[11\coth^2\left(\frac{\pi L}{\lambda}q\right)\operatorname{csch}^4\left(\frac{\pi L}{\lambda}q\right)\right.\right.$$

$$+ 2\coth^4\left(\frac{\pi L}{\lambda}q\right)\operatorname{csch}^2\left(\frac{\pi L}{\lambda}q\right) + 2\operatorname{csch}^6\left(\frac{\pi L}{\lambda}q\right)\Big]\Big]\Big) + \cdots . \tag{33}$$

Expression (33) is especially useful in the regime ($\lambda/L \ll 1$), which is customarily known as the asymptotic regime. Here, the sums may be truncated at just the leading rms, with the result

$$(c_v)_{\text{film}} = \frac{2\pi^2 k}{15\lambda^3}\left[\left\{1 + 60\left[1 - 3\left(\frac{\lambda}{2\pi L}\right)\right]e^{-2\pi L/\lambda}\right\} - \frac{25}{14}\pi^2\left(\frac{kT}{mc^2}\right)^2 \times\left\{1 - 168\left[1 - \left(\frac{\lambda}{\pi L}\right) - \frac{1}{5}\left(\frac{\pi L}{\lambda}\right)\right]e^{-2\pi L/\lambda}\right\} + \cdots\right]. \tag{34}$$

For $\lambda/L \to 0$, we recover the bulk behavior as obtained by Pathria and Singh,[12] see Eq. (17). If, on the other hand, we assume that $kT \ll mc^2$, we recover the corresponding expression of Pajkowski and Pathria[8] for a standard phonon gas in a film geometry.

In order to obtain a useful expansion for the complementary regime ($\lambda/L \gg 1$), we expand the hyperbolic functions in (33) as exponential series, such as

$$\coth x = 1 + 2\sum_{k=1}^{\infty} e^{-2kx},$$

etc. We can then sum over q, with the result

$$(\beta u)_{\text{disp}} = -\frac{1}{\lambda^3}\left(\frac{kT}{mc^2}\right)^2\left[\frac{3}{2\pi^2}\zeta(6)\left(\frac{\lambda}{L}\right)^6 - (2\pi)^3\left(\frac{\lambda}{L}\right)\sum_{k=1}^{\infty} k^4\ln(1 - e^{-2\pi kL/\lambda}) - 4\pi^4\sum_{k=1}^{\infty} k^5(e^{2\pi kL/\lambda} - 1)^{-1}\right] + \cdots \tag{35}$$

This intermediate step is necessary as a preparation for the application of the PSF, for otherwise the extension of the summation in (33) to all values of q is problematic at $q = 0$. That expression, as a whole, is well behaved (in fact, the $q = 0$ term corresponds to the bulk result); however, each sum separately diverges, thus making a direct application of the PSF rather questionable. No such difficulty arises when our expression is written in the form (35) and the PSF can be readily applied, with the result

$$\begin{aligned}(\beta u)_{\text{disp}} = {} & \left(\frac{kT}{mc^2}\right)^2\frac{1}{L^3}\left\{\frac{5\pi^4}{126}\left(\frac{L}{\lambda}\right)^3 - \frac{\pi^4}{2}\left(\frac{\lambda}{L}\right)^3\right. \\ & \times\sum_{j=1}^{\infty}\left[11\coth^2\left(\frac{\pi\lambda}{L}j\right)\operatorname{csch}^4\left(\frac{\pi\lambda}{L}j\right) + 2\coth^4\left(\frac{\pi\lambda}{L}j\right)\operatorname{csch}^2\left(\frac{\pi\lambda}{L}j\right) + 2\operatorname{csch}^6\left(\frac{\pi\lambda}{L}j\right)\right] \\ & - 2\pi^3\left(\frac{\lambda}{L}\right)^2\sum_{j=1}^{\infty} j^{-1}\left[\coth^3\left(\frac{\pi\lambda}{L}j\right)\operatorname{csch}^2\left(\frac{\pi\lambda}{L}j\right) + 2\coth\left(\frac{\pi\lambda}{L}j\right)\operatorname{csch}^4\left(\frac{\pi\lambda}{L}j\right)\right] - 2\pi^2\left(\frac{\lambda}{L}\right) \\ & \times\sum_{j=1}^{\infty} j^{-2}\left[\coth^4\left(\frac{\pi\lambda}{L}j\right) + 2\coth^2\left(\frac{\pi\lambda}{L}j\right)\operatorname{csch}^2\left(\frac{\pi\lambda}{L}j\right)\right] - 6\pi\sum_{j=1}^{\infty} j^{-3}\coth\left(\frac{\pi\lambda}{L}j\right)\operatorname{csch}^2\left(\frac{\pi\lambda}{L}j\right) - 6\left(\frac{L}{\lambda}\right) \\ & \left.\times\sum_{j=1}^{\infty} j^{-4}\operatorname{csch}^2\left(\frac{\pi\lambda}{L}j\right) - \frac{6}{\pi}\left(\frac{L}{\lambda}\right)^2\sum_{j=1}^{\infty} j^{-5}\coth\left(\frac{\pi\lambda}{L}j\right)\right\} + \cdots,\end{aligned} \tag{36}$$

which converges rapidly for $\lambda/L \gg 1$. Retaining only the leading exponential terms, we get

$$(c_v)_{\text{film}} = \frac{3k}{\pi\lambda^2 L}\left(\zeta(3) + 2\left[1 + 2\left(\frac{\pi\lambda}{L}\right) + 2\left(\frac{\pi\lambda}{L}\right)^2 + \frac{4}{3}\left(\frac{\pi\lambda}{L}\right)^3\right]e^{-2\pi\lambda/L} - 10\left(\frac{kT}{mc^2}\right)^2\left\{\zeta(5) + 2\left[1 + 2\left(\frac{\pi\lambda}{L}\right) + 2\left(\frac{\pi\lambda}{L}\right)^2 + \frac{4}{3}\left(\frac{\pi\lambda}{L}\right)^3 + \frac{2}{3}\left(\frac{\pi\lambda}{L}\right)^4 + \frac{4}{15}\left(\frac{\pi\lambda}{L}\right)^5 + \frac{2}{15}\left(\frac{\pi\lambda}{L}\right)^6\right]e^{-2\pi\lambda/L}\right\} + \cdots\right). \tag{37}$$

For $kT \ll mc^2$, this expression reduces to the corresponding one obtained by Pajkowski and Pathria[8] for a phonon gas in a film geometry. On the other hand, if $\lambda/L \to \infty$, the system behaves as if it were a two-dimensional bulk system, with specific heat *per unit area* given by

$$\frac{3k}{\pi\lambda^2}\left[\zeta(3) - 10\zeta(5)\left(\frac{kT}{mc^2}\right)^2 + \cdots\right], \tag{37a}$$

which is indeed expected.

Figure 1 shows the ratio of the actual specific heat, as given by Eqs. (34) and (37), to that of an assembly of nondispersive phonons in the form of a film. It may be mentioned here that although the expressions (34) and (37) are supposed to apply only towards the extreme ends of the range of λ/L, the convergence of the sums involved in the parent expressions (33) and (36) is so rapid that, in practice, we encounter an overlapping range of validity such that, taken together, these expressions constitute an almost complete description of the specific heat of the system. This overlap is rather important because for typical values of L and T, namely 10–100 Å and 0.2–0.5 K, the parameter λ/L indeed lies in the intermediate range, with values of order unity.

It will be noted that in Fig. 1 we have only plotted c_v down to $\lambda/L = 0.4$, despite the fact that the relevant approximation

(34) becomes better and better as we proceed in the direction of decreasing λ/L. The reason for this is that we must as well satisfy the criterion of weak dispersion, viz., $x = mc^2/kT = (mcL/\hbar)(\lambda/L) \gg 1$, which has been assumed in dropping the subsequent terms of (34). Using the mass of an ^{4}He atom and $c = 238$ m/sec, this implies that, for a 25 Å film, we must have $\lambda/L \gg 0.03$. This restriction is not, in fact, serious because we are operating at sufficiently low temperatures anyway.

V. DISPERSIVE EFFECTS IN A CHANNEL GEOMETRY ($\infty \times L \times L$)

In the limit of a channel geometry, the summations appearing in (32) become two-dimensional, with the result

$$(\beta u)_{\text{disp}} = -\frac{1}{\lambda^3}\left(\frac{kT}{mc^2}\right)^2\left(\frac{3}{4\pi^2}\left(\frac{\lambda}{L}\right)^6\theta(3) + \sum_{q_{1,2}=-\infty}^{\infty}{}' \left\{\pi^3\left(\frac{\lambda}{L}\right)q^{-1}\left[\coth^3\left(\frac{\pi L}{\lambda}q\right)\right.\right.\right.$$
$$\times \operatorname{csch}^2\left(\frac{\pi L}{\lambda}q\right) + 2\coth\left(\frac{\pi L}{\lambda}q\right)\operatorname{csch}^4\left(\frac{\pi L}{\lambda}q\right)\Big] - \frac{\pi^4}{4}\Big[11\coth^2\left(\frac{\pi L}{\lambda}q\right)$$
$$\left.\left.\left.\times \operatorname{csch}^4\left(\frac{\pi L}{\lambda}q\right) + 2\coth^4\left(\frac{\pi L}{\lambda}q\right)\operatorname{csch}^2\left(\frac{\pi L}{\lambda}q\right) + 2\operatorname{csch}^6\left(\frac{\pi L}{\lambda}q\right)\Big]\Big]\right) + \cdots, \tag{38}$$

where $q = q_1^2(q_2^2)^{1/2}$ and

$$\theta(s) = \sum_{q_{1,2}=-\infty}^{\infty}{}' (1/q^s) = 4\zeta(s)\beta(s), \tag{39}$$

$\theta(s)$ being the Hardy sums and $\beta(s)$ the Dirichlet L-series $\Sigma_{n=0}^{\infty}(-1)^n(2n+1)^{-s}$. Again, retaining only the leading exponential terms in these summations, we get

$$(c_v)_{\text{channel}} = \frac{2\pi^2 k}{15\lambda^3}\left(\left\{1 + 120\left[1 - 3\left(\frac{\lambda}{2\pi L}\right)\right]e^{-2\pi L/\lambda}\right\} - \frac{25}{14}\pi^2\left(\frac{kT}{mc^2}\right)^2\right.$$
$$\left.\times\left\{1 - 336\left[1 - \left(\frac{\lambda}{\pi L}\right) - \frac{1}{5}\left(\frac{\pi L}{\lambda}\right)\right]e^{-2\pi L/\lambda}\right\} + \cdots\right) \qquad (\lambda/L \ll 1). \tag{40}$$

Comparing this with (34), we find that the finite size correction due to dispersion in the case of a channel geometry is twice that in the case of a film. This was also found to be the case for finite size effects in a standard phonon gas.[9] We can see that this phenomenon is directly related to the dimensionality of the summations in Eqs. (33) and (38); in fact, we can argue that in a cube the finite size correction due to dispersion will be a factor of three larger than in a film. This indeed bears out the intuitive notion that the greater the degree of finiteness of the system, the larger the finite size correction.

To examine the complementary regime ($\lambda/L \gg 1$) for the channel, we transform (38) by applying the two-dimensional PSF to obtain a rapidly convergent result. This formula is now embodied in the identities

$$\sum_{n=1}^{\infty}\frac{1}{(n^2+a^2)} = \frac{\pi}{2a}\coth(\pi a) - \frac{1}{2a^2} \tag{41}$$

and

$$\sum_{q_{1,2}=-\infty}^{\infty}{}'\left(\frac{q}{\mu}\right)K_1(\mu q) = \frac{4\pi}{\mu^4} - \frac{1}{\mu^2} + 4\pi\sum_{q_{1,2}=-\infty}^{\infty}{}'\frac{1}{(\mu^2 + 4\pi^2 q^2)^2}. \tag{42}$$

These identities, along with their relevant derivatives, can be used to re-express (38) after a somewhat tedious but straightforward calculation as

$$(\beta u)_{\text{disp}} = \frac{1}{\lambda^3}\left(\frac{kT}{mc^2}\right)^2\left[\frac{5\pi^4}{126} + \frac{11}{126}\pi^3\left(\frac{\lambda}{L}\right)^2 + 70\left(\frac{\lambda}{L}\right)^3\sum_{n=1}^{\infty}\sum_{q_{1,2}=-\infty}^{\infty}{}'\frac{q^2}{n^3}K_2(2\pi nq\lambda/L)\right.$$
$$- \frac{191}{4}\pi\left(\frac{\lambda}{L}\right)^4\sum_{n=1}^{\infty}\sum_{q_{1,2}=-\infty}^{\infty}{}'\frac{q^3}{n^2}K_3(2\pi nq\lambda/L) - 3\pi^2\left(\frac{\lambda}{L}\right)^5\sum_{n=1}^{\infty}\sum_{q_{1,2}=-\infty}^{\infty}{}'\frac{q^4}{n}K_4(2\pi nq\lambda/L)$$
$$\left. + \frac{5}{4}\pi^3\left(\frac{\lambda}{L}\right)^6\sum_{n=1}^{\infty}\sum_{q_{1,2}=-\infty}^{\infty}{}' q^5 K_5(2\pi nq\lambda/L)\right]. \tag{43}$$

Due to the presence of the modified Bessel functions $K_\nu(z)$, these sums converge rapidly when $\lambda/L \gg 1$. Retaining the requisite number of terms in the asymptotic expansions of the $K_\nu(z)$, we finally obtain

$$(c_v)_{\text{channel}} = \frac{\pi k}{3\lambda L^2}\left(\left\{1 + 48\pi\left(\frac{\lambda}{L}\right)^{5/2}\left[1 + \frac{11}{8}\left(\frac{L}{2\pi\lambda}\right) + \frac{225}{128}\left(\frac{L}{2\pi\lambda}\right)^2\right.\right.$$
$$\left. + \frac{105}{128}\left(\frac{L}{2\pi\lambda}\right)^3 - \frac{525}{512}\left(\frac{L}{2\pi\lambda}\right)^4\right]e^{-2\pi\lambda/L}\Big\} - \frac{3\pi^2}{10}\left(\frac{kT}{mc^2}\right)^2\Big\{1 + 160\pi^2\left(\frac{\lambda}{L}\right)^{11/2}$$
$$\times\left[1 + \frac{7}{8}\left(\frac{L}{2\pi\lambda}\right) + \frac{697}{128}\left(\frac{L}{2\pi\lambda}\right)^2 + \frac{18621}{1024}\left(\frac{L}{2\pi\lambda}\right)^3 + \frac{1431675}{32768}\left(\frac{L}{2\pi\lambda}\right)^4\right.$$
$$\left.\left. + \frac{3686445}{65536}\left(\frac{L}{2\pi\lambda}\right)^5\right]e^{-2\pi\lambda/L}\Big\} + \cdots\Big). \tag{44}$$

For $kT \ll mc^2$, expressions (40) and (44) again reduce to the corresponding ones obtained by Pajkowski and Pathria[8] for a phonon gas in a channel geometry. For $\lambda/L \to 0$, expression (40) reproduces exactly the three-dimensional bulk result (17); however, in the other extreme (when $\lambda/L \to \infty$), expression (44) shows that our system behaves as if it were a one-dimensional bulk system, with specific heat *per unit length* given by

$$\frac{\pi k}{3\lambda}\left[1 - \frac{3\pi^2}{10}\left(\frac{kT}{mc^2}\right)^2 + \cdots\right], \tag{44a}$$

which is indeed expected. Note that the bulk expressions (17), (37a), and (44a) are special cases of the general d-dimensional result, viz.,

$$\frac{k}{2^{d-1}\pi^{d/2}((d/2)-1)!\lambda^d}\left[(d+1)!\zeta(d+1) - \frac{d+2}{8}(d+3)!\,\zeta(d+3)\left(\frac{kT}{mc^2}\right)^2 + \cdots\right], \tag{45}$$

with $d = 3$, 2, and 1, respectively.

Figure 2 shows the ratio of the actual specific heat, as given by Eqs. (40) and (44), to that of an assembly of nondispersive phonons in the form of a square channel.

VI. CONCLUDING REMARKS

As can be seen from the figures, significant deviations from the standard phonon behavior develop when the mean thermal wavelength λ is of the order of the system dimension L. This is quite a reasonable expectation on physical grounds, for dispersion is due in part to multiphonon processes and one expects these processes to be inhibited if the system size is much smaller than the mean thermal wavelength of the phonons. For a typical value of L, say 25 Å, the foregoing condition implies that $T \lesssim 0.7$ K, which is precisely the regime in which the thermodynamic behavior of liquid ^{4}He is dominated by phonons.

Unfortunately we do not have at this stage sufficient experimental data on the low-temperature specific heat of liquid ^{4}He in restricted geometries to offer a meaningful comparison with our calculations. In addition, the problem of the quantitative determination of the intersticial geometry in customarily used media must be improved further so that reliable estimates could be made of the finite size effects expected on theoretical grounds. We hope that the calculations presented here will prompt further experimental work in the directions indicated.

In conclusion we wish to remark that the boundary conditions employed in this paper are only one of the many that can be imposed on the system. In the absence of an *a priori* reason to choose any particular boundary conditions in preference to others, we decided to analyze the problem under periodic boundary conditions which are indeed the simplest to work with; any other choice would have made the mathematics even more cumbersome. Our previous experience with such investigations tells us that while the results obtained by using Dirichlet or Neumann boundary conditions may differ in some important respects from the ones obtained under periodic boundary conditions, the use of the latter does provide us with the first explicit glimpses of the situation expected to arise in a finite system. This would not be the case if one approached the problem by simply adopting a density-of-states appropriate to the given boundary conditions and carrying out integrations, instead of summations, over states. In that case the periodic boundary conditions would not give any finite size effects whatsoever whereas Dirichlet and Neumann boundary conditions would.[17] Such an approach leaves one with the (erroneous) impression that the periodic boundary conditions are rather unreal-

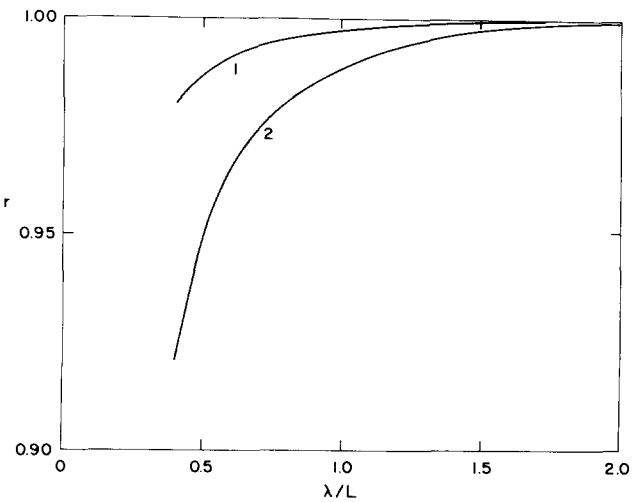

FIG. 1. Ratio of the low-temperature specific heat of a weakly interacting Bose gas to that of a gas of nondispersive phonons, as a function of λ/L, for a film of thickness L. Curve 1: $L = 50$ Å; Curve 2: $L = 25$ Å.

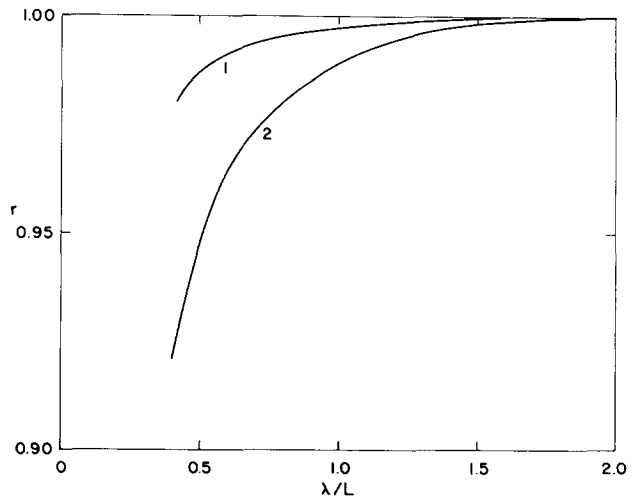

FIG. 2. Same as in Fig. 1, for a square channel of cross section $L \times L$. Curve 1: $L = 50$ Å; Curve 2: $L = 25$ Å.

istic and, at best, an artifice to circumvent the presence of boundaries. Our work over the past few years has, however, shown very clearly that if the summations over states appearing in the expressions for the various physical quantities pertaining to the system are evaluated rigorously, then periodic, as well as Dirichlet and Neumann, boundary conditions lead to finite size effects which are quantitatively determined by the surface-to-volume ratio of the system. This provides the essential motivation with which the calculations reported in this paper were carried out; it also indicates the spirit in which these calculations should be viewed. Of course, if there exists an *a priori* reason to employ a particular set of boundary conditions in preference to others, one is well advised to do so right in the beginning.

In passing, it appears worthwhile to note that the finite size effects reported here should be distinguished from the effects arising explicitly from the surface modes, if any, of the given system. The former are a consequence of the discreteness of the bulk modes which in turn owes to the finiteness of the enclosure; they depend rather strongly on the choice of the boundary conditions imposed on the wavefunctions and do not depend critically on the interparticle interactions. The latter, on the other hand, are a (true) surface phenomenon associated with interfacial tension; they depend rather strongly on the interparticle interactions and do not depend critically on the boundary conditions. Whether one type of effects are more important than the other depends on the system under study and on the properties under investigation.

[1]R.K. Pathria, Phys. Rev. A **5**, 1451 (1972).
[2]S. Greenspoon and R.K. Pathria, Phys. Rev. A **8**, 2657 (1973); **9**, 2103 (1974); **11**, 1080 (1975).
[3]C.S. Zasada and R.K. Pathria, Phys. Rev. A **14**, 1269 (1976); **15**, 2439 (1977).
[4]H.R. Pajkowski and R.K. Pathria, J. Phys. A **10**, 561 (1977).
[5]R.M. Ziff, G.E. Uhlenbeck, and M. Kac, Phys. Lett. C **32**, 169 (1977).
[6]T.C. Padmore, Phys. Rev. Lett. **28**, 1512 (1972).
[7]H. Haug, J. Low Temp. Phys. **12**, 479 (1973).
[8]H.R. Pajkowski and R.K. Pathria, J. Low Temp. Phys. **30**, 697 (1978); **35**, 349 (1979).
[9]C.S. Zasada and R.K. Pathria, J. Phys. A (to be published).
[10]K.A. Brueckner and K. Sawada, Phys. Rev. **106**, 1128 (1957).
[11]R.K. Pathria, *Statistical Mechanics* (Pergamon, New York, 1972), Sec. 7.5.
[12]R.K. Pathria and A.D. Singh, Proc. Nat. Inst. Sci. India A **26**, 520 (1960).
[13]C.S. Meijer, Proc. London Math. Soc. **40**, 1 (1936).
[14]T.D. Lee and C.N. Yang, Phys. Rev. **112**, 1419 (1958).
[15]C.S. Zasada and R.K. Pathria, Phys. Rev. Lett. **29**, 988 (1972); Phys. Rev. A **9**, 560 (1974).
[16]A.N. Chaba and R.K. Pathria, J. Math. Phys. **16**, 1457 (1975); J. Phys. A **9**, 1411, 1801 (1976); **10**, 1823 (1977).
[17]R.K. Pathria, Nuovo Cimento Suppl. **4**, 276 (1966).

The non-Markovian relaxation process as a "contraction" of a multidimensional one of Markovian type

Mauro Ferrario and Paolo Grigolini[a)]

Gruppo Nazionale di Struttura della Materia del C.N.R. and Laboratorio di Chimica Quantistica ed Energetica Molecolare del C.N.R., Istituto di Fisica, Piazza Torricelli 2, 56100 Pisa, Italy

(Received 20 November 1978)

A new approach for obtaining the Fokker–Planck equation to be associated with the generalized Langevin equation is discussed. By using the Mori expansion of the "memory kernel," it is shown that any information of interest may be provided by a suitable multidimensional Fokker–Planck equation of Markovian type. A suitable "contraction" process, furthermore, enables us to find the same two-point conditional probability as the one recently obtained by Fox. This approach may be useful to overcome the Markov approximation which is present in the stochastic Liouville equation theory.

I. INTRODUCTION

The generalized Langevin equation[1] in the monodimensional case reads

$$\frac{du}{dt} = -\int_0^t \beta(t-s)\,u(s)\,ds + \tilde{f}(t), \tag{1}$$

where $\tilde{f}(t)$ is related to the fluctuating force $\tilde{f}'(t)$ by: $\tilde{f}(t) = \tilde{f}'(t)/m$. As usual, we assume that $\langle \tilde{f}(t) \rangle = 0$ and

$$\langle \tilde{f}(t)\tilde{f}(s) \rangle = \Delta^2 \beta(|t-s|) \quad (\Delta^2 = \kappa_B T/m). \tag{2}$$

The possibility of obtaining a Fokker–Planck equation to be associated with Eq. (1) has been the subject of recent investigations.[2-4] Adelman[2] succeeded in deriving from the generalized Langevin equation[1] a sort of generalized Fokker–Planck equation under the general assumption of Gaussian stochastic forces. His equation looks like a usual Fokker–Planck equation but with time dependent friction coefficients. Similar results were obtained by Fox by using a different approach.[3]

More recently,[4] Fox built up in a rigorous way the two-point distributions to be associated with the generalized Langevin equation. Such a distribution was shown to satisfy a Fokker–Planck-like equation which exhibits a two-time dependence and does not appear to be endowed with the nonstationary, Gaussian, Markov nature of the one previously found. The only feature shared with the Markovian processes has to be seen in that both the two-point distribution and the higher order ones depend upon a single two-time function, the velocity autocorrelation

$$\{\langle u(t_1)\,u(t_2)\rangle\} = (k_B T/m)\chi(|t_1 - t_2|). \tag{3}$$

In the present paper we shall show that it is possible to replace the monodimensional form of Eq. (1) with a multidimensional Langevin equation of Markovian type in such a way that the autocorrelation function involving the component of interest of the multidimensional vector is just the autocorrelation function of Eq. (3). The multidimensional Langevin equation we shall find is of the same kind as the one studied by Fox and Uhlenbeck[5] some years ago. As a consequence, both the problem of finding its bona fide Fokker–Planck equation and the one of evaluating the time evolution of the two-point multidimensional distribution does not present any difficulty provided that the Gaussian assumption is performed. It is very interesting to remark that the two-point monodimensional distribution found by Fox can be derived from the multidimensional one through a suitable "contraction" process. It appears, thus, that whereas Fox is able to express everything by the two-time function $\chi_2(t_1 - t_2)$, Eq. (3), we can express any relevant physical property by using both an approximate expression for χ_2 and our multidimensional two-point distribution. A significant advantage of the present theory may be seen in that it affords a straightforward method of evaluating even the relaxation processes of "multiplicative"[6] type which depend on fluctuations of non-Markovian nature.

The leading idea of the present paper is the one already outlined in Ref. 7. According to Wang and Uhlenbeck,[7] in fact, any non-Markovian process can be considered as a "projection" of a more complicated Markov process. However, it is only the more recent theory by Mori[8] which makes it possible to actually obtain the main results of the present paper.

II. THE MULTIDIMENSIONAL LANGEVIN EQUATION OF MARKOV TYPE

According to Mori,[8,9] the Laplace transform of the "memory kernel" $\beta(t)$, Eq. (1), can be written as follows:

$$\hat{\beta}(p) = \cfrac{\Lambda_{n-1}^2}{p + \cfrac{\Lambda_{n-2}^2}{p + \cfrac{\Lambda_{n-3}^2}{p + \cdots \cfrac{\Lambda_2^2}{p + \cfrac{\Lambda_1^2}{p + \beta^{(n)}(p)}}}}}. \tag{4}$$

[a)]To whom any correspondence should be addressed.

 0022-2488/79/122567-06$01.00

We shall show below that truncating such a continued fraction with the assumption that $\hat{\beta}^{(n)}(p)$ be independent of $p(\hat{\beta}^{(n)}(p) = \gamma)$, as suggested by Mori,[8] is equivalent to replacing the monodimensional Langevin equation of non-Markovian kind, Eq. (1), with the following multidimensional Langevin equation:

$$\frac{d}{dt}\mathbf{v} + \mathbf{A}\mathbf{v} + \mathbf{S}\mathbf{v} = \tilde{\mathbf{F}}(t), \tag{5}$$

where $\mathbf{A} =$

$$\begin{pmatrix} 0 & \omega_{12} & 0 & 0 & \cdots & & 0 \\ \omega_{21} & 0 & \omega_{23} & 0 & \cdots & & 0 \\ 0 & \omega_{32} & 0 & \omega_{34} & \cdots & & 0 \\ & & & & & & \\ 0 & 0 & 0 & \cdots & \omega_{n-1\,n-2} & 0 & \omega_{n-1\,n} \\ 0 & 0 & 0 & \cdots & 0 & \omega_{n\,n-1} & 0 \end{pmatrix} \tag{6}$$

and

$$\mathbf{S} = \begin{pmatrix} \gamma & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{7}$$

The vectors $\mathbf{v}$ and $\tilde{\mathbf{F}}$ are column matrices defined as follows: $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\tilde{\mathbf{F}}(t) = (\tilde{F}_1(t), 0, \ldots, 0)$. The constant γ appearing in Eq. (7) is related to $\tilde{F}_1(t)$ by

$$\langle \tilde{F}_1(t)\, \tilde{F}_1(s) \rangle = 2\gamma \Delta^2 \delta(t-s), \tag{8}$$

in such a way as to satisfy the "fluctuation–dissipation" relation

$$\langle \tilde{\mathbf{F}}(t)\, \tilde{\mathbf{F}}(s) \rangle = 2\Delta^2 \mathbf{S}\delta(t-s). \tag{9}$$

By now we have used a kind of averaging with respect to the stochastic force, denoted with the symbol $\langle \cdots \rangle$, and one with respect to the initial velocity of interest, $v_n(0)$, denoted with the symbol $\{\cdots\}$. We have followed the same notations as the ones introduced by Fox[4] according to Ornstein and Uhlenbeck.[10] In the following we shall need also a new kind of averaging with respect to the initial velocities of noninterest. Such a new kind of average will be denoted with the symbol $[\cdots]$. Any velocity distribution will be assumed to be Maxwellian.

In order to demonstrate the equivalence between the monodimensional generalized Langevin equation and the multidimensional one of Markovian type, we shall follow an induction method. For $n = 1$ the usual Langevin equation may be written in the following way:

$$\dot{v}_1(t) = -\int_0^t \beta_1(t-\tau)\, v_1(\tau)\, d\tau + \tilde{f}_1(t), \tag{10}$$

where the Laplace transform of β_1 is given by

$$\hat{\beta}_1^{(n)}(p) = \gamma. \tag{11}$$

With the symbol $\hat{\beta}_n(p)$, $n = 1,2,\cdots$, we denote the approximate value of $\hat{\beta}(p)$, Eq. (4), obtained by replacing $\hat{\beta}^{(n)}(p)$ with the constant value γ. In the $(n-1)$-dimension case we assume that the variable of interest, $v_{n-1}(t)$, satisfies the generalized Langevin equation

$$\dot{v}_{n-1} = -\int_0^t \beta_{n-1}(t-\tau) v_{n-1}(\tau)\, d\tau + \tilde{f}_{n-1}(t). \tag{12}$$

We have now to show that such a property is true also in the n-dimension case. For we assume the theorem be true in the $(n-1)$-dimension case, the two last components of $\mathbf{v}$ can be given the following expression:

$$\dot{v}_{n-1} = -\int_0^t \beta_{n-1}(t-\tau) v_{n-1}(\tau)\, d\tau - \omega_{n-1\,n} v_n + \tilde{f}_{n-1}(t), \tag{13}$$

$$\dot{v}_n = -\omega_{n\,n-1} v_{n-1}. \tag{13'}$$

Performing the Laplace transform of such a system and eliminating the variable $\hat{v}_{n-1}(p)$, we obtain

$$-v_n(0) + \left(p - \frac{\omega_{n\,n-1}\omega_{n-1\,n}}{p + \hat{\beta}_{n-1}(p)}\right)\hat{v}_n(p) = -\frac{\omega_{n\,n-1}}{p + \hat{\beta}_{n-1}(p)}\left[v_{n-1}(0) + \hat{f}_{n-1}(p)\right]. \tag{14}$$

We can identify the previous equation in the Laplace transform of the nth order aproximation to the generalized Langevin equation, provided that

$$\hat{\beta}_n(p) = -\frac{\omega_{n\,n-1}\omega_{n-1\,n}}{p + \hat{\beta}_{n-1}(p)} \tag{15}$$

and

$$\hat{f}_n(p) = -\frac{\omega_{n\,n-1}}{p + \hat{\beta}_{n-1}(p)}\left[v_{n-1}(0) + \hat{f}_{n-1}(p)\right]. \tag{16}$$

From a comparison with Eq. (4) we also obtain

$$-\omega_{n\,n-1}\omega_{n-1\,n} = \Lambda^2_{n-1}. \tag{17}$$

If we also assume that

$$\omega_{n-1\,n} = -\omega_{n\,n-1}, \tag{18}$$

we obtain as many independent parameters as Mori mathematical constraints. The matrix A is, furthermore, given a nondissipative character. In Appendix A for any n-dimension case we shall show that also the constraint involved by the fluctuation–dissipation theorem, Eq. (2), is satisfied, i.e., we shall obtain

$$[\langle \tilde{f}_n(t)\, \tilde{f}_n(s) \rangle] = \Delta^2 \beta_n(|t-s|). \tag{19}$$

We have thus demonstrated that, according to the idea outlined by Wang and Uhlenbeck[7] the expression of β in terms of continued fractions is equivalent to regarding the non-Markovian process expressed by Eq. (1) as a "contraction" of a higher dimension Markov process. We would like to stress that the variables $v_1, v_2, \ldots, v_{n-1}$ can be regarded as a sort of simplified or "reduced" thermal bath[11] for our variable of interest. In Sec. IV, in fact, we shall show that any order of the Mori approximation can be given in interpretation in terms of physical models. Though the replacement of the generalized Langevin equation with a Markovian one of finite dimensions involve an unavoidable approximation, the general demonstration of this section makes it clear that it can be accomplished within any desired degree of accuracy.

III. THE MULTIDIMENSIONAL FOKKER-PLANCK EQUATION

The Fokker–Planck equation concerning the multidimensional Markovian Langevin equation (5) can easily be found by noting that Eq. (5) is of the same kind as the one studied by Fox and Uhlenbeck.[5] By taking into account that it is just our simple matrix S which enters the fluctuation–dissipation theorem, we obtain that Eq. (40) of Sec. I2 of Ref. 12 directly provides for the multidimensional probability, $P_2^{(M)}(\mathbf{v}(1)t_1;\mathbf{v}(2)t_2)$ the following Fokker–Planck equation

$$\frac{\partial P_2^{(M)}}{\partial t} = \sum_{ij} \frac{\partial}{\partial v_i(2)} A_{ij} v_j(2) P_2^{(M)} + \gamma \frac{\partial}{\partial v_1(2)} \left(v_1(2) + \Delta^2 \frac{\partial}{\partial v_1(2)} \right) P_2^{(M)}. \tag{20}$$

Following Ref. 12, we could obtain Eq. (20) under the important assumption of stochastic Gaussian forces. In Appendix B we shall show that the multidimensional probability $P_2^{(M)}$ when suitably "contracted" becomes the two-point conditional distribution recently found by Fox.[4] Such a result is in full agreement with the physical requirement that the generalized Langevin equation (1) univocally define its associated Fokker–Planck equation.

IV. THE MORI TRUNCATION INTERPRETED IN TERMS OF PHYSICAL MODELS

Any truncation of the Mori continued fraction can be given a physical interpretation through models. The first nontrivial truncation ($n = 2$), for example, provides just the stochastic harmonic oscillator studied by Fox and Uhlenbeck[5] [see their Eqs. (32) and (33)]. In such a case the variable of interest, apart from a constant factor, can be identified in the position of the harmonic oscillator.

As far as the truncation at $n = 3$ is concerned, we can show that the itinerant harmonic oscillator[13a,14] can be regarded as its "reduced" physical model. Such a physical model consists of a disk coupled through an elastic force with an external anulus which is the only component of the two-body system undergoing a fluctuation–dissipation process. The corresponding Langevin equation can be written as follows:

$$\dot{\Omega} = \ddot{\theta} = -\Lambda_2^2(\theta - \psi), \tag{21}$$

$$\ddot{\psi} = \Lambda_1^2(\theta - \psi) - \gamma\dot{\psi} + \tilde{\lambda}(t). \tag{21'}$$

θ is the angular coordinate of the disk and ψ is the angular coordinate of the anulus. By identifying $\dot{\theta}$ with v_3 (the variable of interest), $\Lambda_2(\theta - \psi)$ with v_2, $(\Lambda_2/\Lambda_1)\dot{\psi}$ with v_1 and $\Lambda_2\tilde{\lambda}(t)/\Lambda_1$ with $\tilde{F}_1(t)$, we can replace the system of Eq. (21) with our multidimensional Langevin equation truncated at $n = 3$. We can also follow a direct approach. By performing the Laplace transform of Eqs. (21) and eliminating the irrelevant variables, one can directly obtain the Laplace transform of Eq. (1) with β truncated at $n = 3$. It is interesting to notice that any odd truncation can be interpreted in terms of a generalized itinerant oscillator. In fact, one can easily show that adding a further anulus to the two-component itinerant oscillator corresponds to the Mori truncation at $n = 5$, and so on. The itinerant oscillator as a physical model corresponding to the Mori truncation at $n = 3$ has already been discussed by Davies *et al.*[13b]

V. CONCLUDING REMARKS

We have solved the problem of finding the Fokker–Planck equation to be associated with the generalized Langevin equation in an alternative way with respect to the approach of both Adelman[2] and Fox.[3,4] Furthermore, our theory could allow us to eliminate also the assumption of Gaussian stochastic forces. We could indeed express everything by using our multidimensional probability $P_2^{(M)}$ provided that a Markov Fokker–Planck equation is defined in the absence of Gaussianness. As far as the non-Markovian nature of the Fokker–Planck-like equation found by Fox[4] is concerned, it can be easily reproduced in the context of the present approach by noticing that the Smoluchowski equation is affected in form by the "contraction" process to be applied in order to eliminate the variables of noninterest.

A significant advantage provided by our approach has to be seen in the possibility of exploiting the well-known stochastic Liouville equation theory[15,16] in such a way as to include non-Markov stochastic processes. Consider, in fact, the equation

$$\dot{u} = F(u,t,y(t)) \tag{22}$$

where $y(t)$ is a stochastic variable obeying the generalized Langevin equation. The theory developed in the present paper allows us to build up an enlarged set of stochastic variables, $\{\lambda\}$, including $y(t)$, in such a way that the "contracted" dynamics only involving $y(t)$ is the same as the one provided by Eq. (1). However, the probability density $\Pi(\{\lambda\},t)$ concerning the enlarged set of variables obeys the Markov master equation

$$\dot{\Pi}(\{\lambda\},t) = W\Pi(\{\lambda\},t), \tag{23}$$

where W is our multidimensional Markoff diffusion operator. By following the basic idea of the stochastic Liouville equation theory,[15,16] we notice that the multidimensional variable $(u,\{\lambda\})$ is again a Markov process which according to Ref. 16 obeys the following master equation:

$$\frac{\partial}{\partial t}\mathscr{P}(u,\{\lambda\},t) = -\frac{\partial}{\partial u}[F(u,y)\mathscr{P}] + W\mathscr{P}. \tag{24}$$

Equation (24) can be regarded as a generalized stochastic Liouville equation avoiding the approximation of Markov relaxation over the stochastic variable y.

The possibility of using a Markov diffusion process like the one of Eq. (24) has already been stressed by Kubo.[17] He asserted,[17] in fact, that any stochastic variable y may be supplemented with additional variables to form a complete set of Markovian variables. The theory of this paper shows how such a task may actually be accomplished.

The initial motivation for the present study has to be looked for in our desire that the "reduced" model theory[11] may be given a solid support. Though the present investigation is classical in nature, its extension to a quantum-mechanical formalism does not involve, in our opinion, any

difficulty. It seems, thus, that the molecular radiationless theory, which is the subject dealt with in Ref. 11, may be formulated in a very rigorous fashion along the lines illustrated in the present paper.

It should be emphasized, furthermore, that any finding of the present paper was allowed for by an intensive exploitation of the results reported in the excellent review by Fox[12] as well as in his recent papers.[3,4]

APPENDIX A: PROOF OF THE FLUCTUATION-DISSIPATION THEOREM[19]

We recall an important result reported by Fox[4] about the double Laplace transform, which can be written as follows:

$$\int_0^\infty dt_2 \int_0^\infty dt_1\, e^{-pt_2} e^{-p't_1} f(|t_1 - t_2|)$$

$$= [\hat{f}(p) + \hat{f}(p')]/(p + p'). \tag{A1}$$

For $n = 1$ the fluctuation–dissipation theorem reads

$$\langle \tilde{f}_1(t_1)\tilde{f}_1(t_2)\rangle = 2\Delta_1^2 \gamma \delta(|t_1 - t_2|) = \Delta_1^2 \beta_1(|t_1 - t_2|). \tag{A2}$$

We assume the theorem be satisfied in the $(n-1)$-dimensional case and be written as

$$[\langle \tilde{f}_{n-1}(t_1)\tilde{f}_{n-1}(t_2)\rangle] = \Delta_{n-1}^2 \beta_{n-1}(|t_1 - t_2|) \tag{A3}$$

By using Eq. (A1) the Laplace transform of Eq. (A3) can be written as

$$[\langle \hat{f}_{n-1}(s)\hat{f}_{n-1}(p)\rangle]$$
$$= \Delta_{n-1}^2 [\hat{\beta}_{n-1}(p) + \hat{\beta}_{n-1}(s)]/(s+p). \tag{A4}$$

For the n-dimension case, by exploiting Eqs. (15) and (16) we can write the double Laplace transform of

$$[\langle \tilde{f}_n(t_1)\tilde{f}_n(t_2)\rangle] \tag{A5}$$

in the following way:

$$[\langle \hat{f}_n(p)\hat{f}_n(s)\rangle] = (\omega_{n-1\,n})^{-2}\beta_n(p)\beta_n(s)$$
$$\times \{[v_{n-1}(0)v_{n-1}(0)]$$
$$+ \langle \hat{f}_{n-1}(s)\hat{f}_{n-1}(p)\rangle\}. \tag{A6}$$

In order to obtain the previous relation, we exploited the kind of averaging denoted by the symbol [···]. By using (A4) we can write

$$[\langle \hat{f}_n(p)\hat{f}_n(s)\rangle]$$
$$= \frac{\Delta_{n-1}^2}{(s+p)}\left(\hat{\beta}_n(p)\hat{\beta}_n(s)\frac{s + \hat{\beta}_{n-1}(s)}{\omega_{n-1\,n}^2} + \hat{\beta}_n(p)\hat{\beta}_n(s)\frac{p + \hat{\beta}_{n-1}(p)}{\omega_{n-1\,n}^2}\right)$$
$$= \Delta_{n-1}^2 \frac{\hat{\beta}_n(p) + \hat{\beta}_n(s)}{(s+p)}. \tag{A7}$$

By applying the inverse double Laplace transform we can write

$$[\langle \tilde{f}_n(t_1)\tilde{f}_n(t_2)\rangle] = \Delta_{n-1}^2 \beta_n(|t_1 - t_2|). \tag{A8}$$

The parameters Δ_i used in the course of this demonstration are defined by

$$\Delta_i^2 = [v_i(0)v_i(0)] \quad (i < n), \tag{A9}$$

$$\Delta_i^2 = \{v_i(0)v_i(0)\} \quad (i = n). \tag{A9'}$$

Equation (A8) results in the desired theorem provided that $\Delta_1 = \Delta_2 = \cdots \Delta_n = \Delta$.

APPENDIX B: PROOF OF THE EQUIVALENCE BETWEEN THE TWO-POINT CONDITIONAL PROBABILITY OBTAINED BY "CONTRACTION" AND THE ONE BY FOX[4]

Even this demonstration will be performed by using a mathematical induction method. The lowest-dimension case is well known. For the $(n-1)$-dimension case we assume that our two-point "contracted" probability, $\mathscr{C}(P_2)$, satisfies

$$\mathscr{C}(P_2^{(M)}({}^{(2)}\mathbf{v}, t_2; {}^{(1)}\mathbf{v}, t_1))$$
$$= p_2({}^{(2)}v_{n-1}, t_2; {}^{(1)}v_{n-1}, t_1)$$
$$= \{[1 - \chi_{n-1}^2(t_2 - t_1)]2\pi\Delta^2\}^{-1/2}$$
$$\times \exp\left(-\frac{1}{2\Delta^2}\frac{[{}^{(2)}v_{n-1} - \chi_{n-1}(t_2 - t_1)\,{}^{(1)}v_{n-1}]^2}{1 - \chi_{n-1}^2(t_2 - t_1)}\right), \tag{B1}$$

where we used the result by Fox.[4] Following his formalism the single Laplace transform of $\chi_{n'}(t)$ is

$$\hat{\chi}_{n'}(p) = [p + \hat{\beta}_{n'}(p)]^{-1}. \tag{B2}$$

In the n-dimension case we can remark that the multidimension Langevin equation (5) can be written as

$$\frac{d}{dt}\begin{pmatrix} v_n \\ v_{n-1}\end{pmatrix} = \begin{pmatrix} 0 & \omega_{n-1\,n} \\ -\omega_{n-1\,n} & 0\end{pmatrix}\begin{pmatrix} v_n \\ v_{n-1}\end{pmatrix}$$
$$- \int_0^t \begin{pmatrix} 0 & 0 \\ 0 & \beta_{n-1}(t-s)\end{pmatrix}$$
$$\times \begin{pmatrix} v_n(s) \\ v_{n-1}(s)\end{pmatrix} ds + \begin{pmatrix} 0 \\ \tilde{f}_{n-1}(t)\end{pmatrix}. \tag{B3}$$

Equation (23) expresses also the generalized Langevin equation for the harmonic oscillator.[3] Fox,[3,12] found the solution of this problem in the following form:

$$P_2({}^{(2)}\mathbf{u}, t_2; {}^{(1)}\mathbf{u}, t_1) = \left(\frac{\|[\mathbf{1} - \mathbf{M}(t_2 - t_1)\mathbf{M}^+(t_2 - t_1)]^{-1}\|}{(2\pi\Delta^2)^2}\right)^{1/2}$$
$$\times \exp\left\{-\frac{1}{2\Delta^2}[{}^{(2)}\mathbf{u} - \mathbf{M}(t_2 - t_1)\,{}^{(1)}\mathbf{u}]^+\right.$$
$$\times [\mathbf{1} - \mathbf{M}(t_2 - t_1)\mathbf{M}^+(t_2 - t_1)]^{-1}$$
$$\left.\times \; [{}^{(2)}\mathbf{u} - \mathbf{M}(t_2 - t_1)\,{}^{(1)}\mathbf{u}]\right\}, \tag{B4}$$

where the vector $\mathbf{u}$ indicates the following bidimensional

vector,

$$\mathbf{u} = (v_n, v_{n-1}), \tag{B5}$$

and the single Laplace transform of the matrix $\mathbf{M}(t)$ is defined as

$$\hat{M}(p) = [p\mathbf{1} - i\Omega + \hat{\Phi}(p)]^{-1}. \tag{B6}$$

The matrix Ω is defined by

$$i\Omega = \begin{pmatrix} 0 & \omega_{n-1\,n} \\ -\omega_{n-1\,n} & 0 \end{pmatrix}. \tag{B7}$$

The matrix $\hat{\Phi}(p)$ is the Laplace transform of the matrix

$$\Phi(t) = \begin{pmatrix} 0 & 0 \\ 0 & \beta_{n-1}(t) \end{pmatrix}. \tag{B8}$$

The matrices Ω and Φ are involved by Eq. (B3). In order to obtain $\mathscr{C}(P_2^{(M)})$ we must use the following identity

$$w_2({}^{(2)}v_n, t_2; {}^{(1)}v_n, t_1) = w_1({}^{(1)}v_n)\, p_2({}^{(2)}v_n, t_2; {}^{(1)}v_n, t_1)$$

$$= \int\int d\,{}^{(2)}v_{n-1}\, d\,{}^{(1)}v_{n-1}\, W_2({}^{(2)}v_n, {}^{(2)}v_{n-1}, t_2; {}^{(1)}v_n, {}^{(1)}v_{n-1}, t_1)$$

$$= \int\int d\,{}^{(2)}v_{n-1}\, d\,{}^{(1)}v_{n-1}\, W_1({}^{(1)}v_n, {}^{(1)}v_{n-1}) P_2({}^{(2)}v_n, {}^{(2)}v_{n-1}, t_2; {}^{(1)}v_n, {}^{(1)}v_{n-1}, t_1). \tag{B9}$$

It is convenient to remark that any probability function concerning Eq. (B3) has been denoted with capital letters. By writing in an explicit way $\mathbf{M}(p)$ and using the identity (15), we can obtain the following expression for $\mathbf{M}(t_1 - t_2)$:

$$\mathbf{M}(t_2 - t_1) = \begin{pmatrix} \chi_n(t_2 - t_1) & -\dfrac{1}{\omega_{n-1\,n}}(\beta_n * \chi_n)(t_2 - t_1) \\ \dfrac{1}{\omega_{n-1\,n}}(\beta_n * \chi_n)(t_2 - t_1) & \dfrac{d}{dt}\left(\dfrac{1}{\omega_{n-1\,n}^2}\right)(\beta_n * \chi_n)(t_2 - t_1) \end{pmatrix}. \tag{B10}$$

The symbol $(\alpha * \beta)(t)$ has the following meaning:

$$(\alpha * \beta)(t) = \int_0^t \alpha(t)\beta(t - \tau)\, d\tau. \tag{B11}$$

We define the matrix $\mathbf{A}(t_2 - t_1)$ through

$$\mathbf{A}^{-1} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} = 1 - \mathbf{M}(t_2 - t_1)\mathbf{M}^*(t_2 - t_1), \tag{B12}$$

where a, b, and c are suitably defined when $\mathbf{M}$ is given its explicit expression, Eq. (B10). $\mathbf{A}$ is the matrix of the coefficients of the quadratic form appearing in the exponential of Eqs. (B4). W_2 is proportional to such an exponential. By integrating W_2, Eq. (B9), we now obtain

$$\int W_2 d\,{}^{(2)}v_{n-1} \propto W_1({}^{(1)}\mathbf{u}) \exp\left[-\frac{1}{2\Delta^2} \Delta v_n^2 \left(\frac{1}{a}\right)\right], \tag{B13}$$

where Δv_n is defined by

$${}^{(2)}\mathbf{u} - \mathbf{M}(t_2 - t_1)\,{}^{(1)}\mathbf{u} = \Delta\mathbf{u} = \begin{pmatrix} \Delta v_n \\ \Delta v_{n-1} \end{pmatrix}. \tag{B14}$$

Up to this point, by writing Eq. (B13) in an explicit way we obtain

$$\int W_2 d\,{}^{(2)}v_{n-1} \propto \exp\left\{ -\frac{1}{2\Delta^2} {}^{(2)}v_n - {}^{(1)}v_n \chi_n(t_2 - t_1) - \frac{1}{\omega_{n-1\,n}}(\beta_n * \chi_n)(t_2 - t_1)\,{}^{(1)}v_{n-1} \right]^2 \times \frac{1}{1 - \chi_n^2(t_2 - t_1) - \Delta\, 1/\omega_{n-1\,n}^2 [(\beta_n * \chi_n)(t_2 - t_1)]^2} \Bigg\}$$

$$\times \exp\left(-\frac{1}{2\Delta^2}({}^{(1)}v_{n-1}^2 + {}^{(1)}v_n^2)\right). \tag{B15}$$

We have now to integrate Eq. (B15) over ${}^{(1)}v_{n-1}$. We may then write

$$\int W_2 d\,{}^{(2)}v_{n-1}\, d\,{}^{(1)}v_{n-1} \propto \int d\,{}^{(1)}v_{n-1} \times \exp\left\{ -\frac{1}{\Delta^2}[{}^{(2)}v_n^2 + {}^{(1)}v_n^2 - 2\,{}^{(2)}v_n\,{}^{(1)}v_n \chi_n(t_2 - t_1)]^2 \times \frac{1}{1 - \chi_n^2(t_2 - t_1)} - \frac{1}{2\Delta^2} \frac{1}{1 - \chi_n^2 - (1/\omega_{n-1\,n}^2)(\beta_n * \chi_n)^2} \times \left[(1 - \chi_n^2)^{1/2}\,{}^{(1)}v_{n-1} - \frac{(\beta_n * \chi_n)}{\alpha_{n-1n}(1 - \chi_n^2)^{1/2}}\,{}^{(2)}v_n + \frac{\chi_n(\beta_n * \chi_n)\,{}^{(1)}v_n}{\omega_{n-1n}(1 - \chi_n^2)^{1/2}} \right]^2 \right\}. \tag{B16}$$

We wrote Eq. (B16) in such a way as to allow for its integration in an easy way. By actually performing such an integra-

tion over ${}^{(1)}v_{n-1}$ and by recalling the relation between w_2 and p_2, whose explicit form is given in Eq. (B9), we obtain

$$p_2({}^{(2)}v_n,t_2;{}^{(1)}v_n,t_1) \propto \exp\left(-\frac{1}{2\Delta^2}\frac{[{}^{(2)}v_n-\chi_n(t_2-t_1)\,{}^{(1)}v_n]^2}{1-\chi_n^2(t_2-t_1)}\right), \tag{B17}$$

which is just the result by Fox, which was assumed to be true in the $(n-1)$-dimension case, Eq. (B1).

[1]H. Mori, Prog. Theor. Phys. **33**, 423 (1965).
[2]S.A. Adelman, J. Chem. Phys. **64**, 124 (1976).
[3]R.F. Fox, J. Stat. Phys. **16**, 259 (1977).
[4]R.F. Fox, J. Math. Phys. Phys. **18**, 2331 (1977).
[5]R.F. Fox and G.E. Uhlenbeck, Phys. Fluids **13**, 1893 (1970).
[6]R.F. Fox, J. Math. Phys. **13**, 1196 (1972).
[7]M.C. Wang and G.E. Uhlenbeck, in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954), p. 115.
[8]H. Mori, Prog. Theor. Phys. **34**, 399 (1965).
[9]B.J. Berne and G.D. Harp, Advan. Chem. Phys. **62**, 63 (1970).
[10]G.E. Uhlenbeck and L.S. Ornstein, Phys. Rev. **36**, 823 (1930).
[11]P. Grigolini, Chem. Phys. Lett. **47**, 483 (1977); P. Grigolini and A. Lami, Chem. Phys. **30**, 61 (1978).
[12]R.H. Fox, "Gaussian Stochastic Processes in Physics," preprint (1978).
[13](a) N.E. Hill, Proc. Phys. Soc. **82**, 723 (1963); (b) A.R. Davies, G.J. Evans, and M.W. Evans, J. Mol. Struct. **46**, 395 (1978).
[14]W.T. Coffey, T. Ambrose, and J.H. Calderwood, J. Phys. D **9**, L115 (1976).
[15]R. Kubo, Advan. Chem. Phys. **15**, 101 (1969).
[16]N.G. van Kampen, Phys. Rep. C **24**, 171 (1976).
[17]R. Kubo, J. Phys. Soc. Japan **26**, 1 (1969).

Linear response theory revisited. II. The master equation approach

K. M. van Vliet

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Quebec H3C 3J7, Canada and Department of Electrical Engineering, University of Florida, Gainesville, Florida 32611

(Received 31 August 1978; accepted for publication 15 February 1979)

We consider a system perturbed by an external field and subject to dissipative processes. From the von Neumann equation for such a system in the weak coupling limit we derive an inhomogeneous master equation, i.e., a master equation with dissipative terms and streaming terms, using Zwanzig's projection operator technique in Liouville space. From this equation the response function, as well as expressions for the generalized conductivity and susceptibility, is obtained. It is shown that for large times only the diagonal part of the density operator is required. The various expressions are found to be in complete harmony with previous results (Part I) obtained via the van Hove limit of the Kubo–Green linear response formulas. In order to account for the properties at quantum frequencies, the evolution of the nondiagonal part in the weak coupling limit is also established. The complete time dependent behavior of the dynamic variables in the van Hove limit is expressed by $B(t) = \exp[-(\Lambda_d - iL^0)t]B$, where Λ_d is the master operator and L^0 the Liouville operator in the interaction picture. The cause of irreversibility is discussed. Finally, the inhomogeneous master equation is employed to obtain as first moment equation a Boltzmann equation with streaming terms, applicable to quantum systems.

1. INTRODUCTION

In a previous paper[1] of the same title, Part I, a reinterpretation of Kubo's linear response theory[2] was given, by considering a system with a partitioned Hamiltonian $H = H^0 + \lambda V$ and by applying the van Hove limit[3] to the Kubo-Green formulas for the various transport functions and transport coefficients. Here H^0 is the Hamiltonian of what we consider the "motion proper" in the system, whereas λV represents an interaction which causes random transitions between the states of H^0, thus giving rise to dissipation. The new expressions showed a clear convergence and approach to equilibrium for $t \to \infty$. The general behavior of the diagonal parts of the Heisenberg operators was found to be of the form $B_d^R(t) = e^{-\Lambda_d t} B_d$, where the superscript R (for reduced operators) indicates the result after the van Hove limit; Λ_d was an operator associated with the transport operator of the master equation. Specifically, we had, with $\{|\gamma\rangle\}$ being the eigenstates and $\{\epsilon_\gamma\}$ being the eigenvalues of H^0,

$$\Lambda_d K_d = -\sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma| [W_{\gamma'\gamma} D_{\gamma'} - W_{\gamma\gamma'} D_\gamma] K_d, \quad (1.1)$$

where

$$D_\gamma K_d = \langle\gamma|K_d|\gamma\rangle, \quad (1.2)$$

and where K_d is an arbitrary diagonal operator[4]; $W_{\gamma\gamma'}$ is the transition probability due to the dissipative processes λV:

$$W_{\gamma\gamma'} = (2\pi\lambda^2/\hbar)|\langle\gamma|V|\gamma''\rangle|^2 \delta(\epsilon_\gamma - \epsilon_{\gamma'}) = W_{\gamma'\gamma}. \quad (1.3)$$

The equality of the first and the third member expresses the property of microscopic reversibility.

The method followed in Part I is indicated by the arrows (a) and (b) of the summarizing scheme of Fig. 1. Starting with the full Liouville or von Neumann equation, the response functions were obtained as in standard linear response theory, arrow (a); then these results were reinterpreted by evaluating the van Hove limit of these expressions, arrow (b). Presently, we will follow a different route. First, from the von Neumann equation we derive, by projection operator techniques, a new inhomogeneous master equation for the density operator or specifically for its diagonal part ρ_d, pertaining to the states of H^0 [see arrow (c)]. This equation differs from the standard master equation in that it has streaming terms relating to the applied external fields. Next, we employ this equation in order to obtain the response and the transport coefficients [arrow (d)]. However, *the new linear response theory, indicated by (d), operates entirely within the subdynamics of* H^0; no reference whatsoever is made to the Heisenberg operators and the full microscopic evolution of the original von Neumann equation. The inhomogeneous master equation contains all the information for the evolution of $\rho_d[H^0]$.

Since we will try to make this article largely self contained, we review some points which were stated in Part I. First, by the van Hove limit for weakly interacting systems we meant

$$\lambda \to 0, \quad t/\tau_t \to \infty, \quad \lambda^2 t \quad \text{finite}. \quad (1.4)$$

This indicates that, when the scaling factor λ of the interaction goes to zero, the time interval over which the system must be considered, scaled with respect to the time for transitions to take place, $\tau_t \approx \hbar/\delta\epsilon$, becomes very large. However, in this limiting process, terms of order $\lambda^2 t$ must be maintained. This is easily seen as follows: The reciprocal time for relaxation of the distribution function is of the order of $W_{\gamma\gamma'}$, i.e., $\tau_r^{-1} = \text{constant} \times \lambda^2$. The time interval over which the distribution is observed, on the other hand, is to be scaled with respect to τ_r, i.e., t/τ_r is bounded, so that $t^{-1} = K \cdot \text{const.} \times \lambda^2$. Thus, $\lambda^2 t$ is finite.

Secondly, we noted in Part I that for asymptotic times all macroscopic operators which are somewhat coarse grained (see Sec. 2 for details) behave as near-diagonal operators, i.e., they (almost) commute with the Hamiltonian H^0.

 0022-2488/79/122573-23$01.00

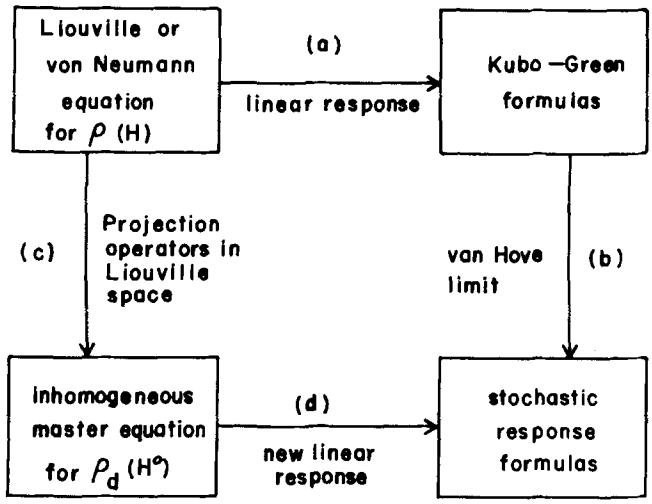

FIG. 1. Scheme of methods employed. Arrows (a) and (b) refer to the derivations of Part I, arrows (c) and (d) to the derivations of this paper.

This presents problems for the response formulas which involve commutators of macroscopic and usually unbounded operators. In Part I this was alleviated by considering mainly the classical frequency limit, defined by

$$\hbar\omega\beta\to 0 \text{ or } t/\hbar\beta\to\infty \quad (\beta = 1/kT). \tag{1.5}$$

In the present paper we will see that the occurrence of these commutators requires a judicial application of the projection operators which select diagonal parts. We will, however, go beyond the limit (1.5) and also investigate the behavior for quantum frequencies. To this end we derive an evolution equation for the nondiagonal part of the density operator ρ_{nd}. It will be shown that for classical frequencies the effect of $\rho_{nd}(t)$ is usually negligible with respect to $\rho_d(t)$. However, for quantum frequencies the nondiagonal parts are not negligible; in particular, these parts are necessary in order to establish the full fluctuation–dissipation theorem, including the quantum correction factor.

This paper is divided as follows: In Sec. 2 we discuss some mathematical preliminaries. In Secs. 3 and 4 the diagonal and nondiagonal evolution equations, are developed and the van Hove limit is carried through. In Sec. 5 we discuss the impact of these derivations for the phenomenon of irreversibility. We show that the assumption that $\rho(0)$ is invariant with respect to translation over the dimensions of the system, together with the application of the van Hove limit to the partitioned Hamiltonian $H = H^0 + \lambda V$, is sufficient to arrive at irreversible, dissipative behavior.

Further, in Sec. 6 we establish the new response formulas for classical frequencies and in Sec. 7 some results for quantum frequencies are discussed.

Finally, in Sec. 8 we deal with another application of the inhomogeneous master equation, in that we derive a Boltzmann equation with streaming terms for quantum systems. Our developments here parallel those of van Hove[5] for the Boltzmann equation without streaming.

2. SOME MATHEMATICAL PRELIMINARIES

A. Diagonal and nondiagonal parts

The state space pertaining to the subdynamics of H^0 be $\mathscr{H}$; the eigenstates of H^0 spanning this space are denoted as $\{|\gamma\rangle\}$. The wave mechanical form of these states will be denoted by $\gamma(\{q_k\}) = \langle\{q_k\}|\gamma\rangle$. For γ we assume periodic boundary conditions over the macroscopic volume of the system, such as is customary in solid state physics for systems capable of carrying steady state currents[6]. We also assume that $\mathscr{H}$ is part of a bigger space $\mathscr{S}$ in which the scalar product is still defined as in $\mathscr{H}$, but without the restriction of periodic boundary conditions for the wave mechanical form $\phi\{q_k\})$ of its elements $|\phi\rangle$. Let K be an arbitrary operator of $\mathscr{S}$. Using the closure property for the states $\{|\gamma\rangle\}$, we can generally write

$$K = \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma|K|\gamma'\rangle\langle\gamma'| = \sum_{\gamma} |\gamma\rangle\langle\gamma|K|\gamma\rangle\langle\gamma| + \sum_{\gamma\neq\gamma'} |\gamma\rangle\langle\gamma|K|\gamma'\rangle\langle\gamma'| = \mathscr{P}K + (1-\mathscr{P})K, \tag{2.1}$$

where the two expressions of the middle member define the diagonal part $K_d = \mathscr{P}K$ and the nondiagonal part $K_{nd} = (1-\mathscr{P})K$. From the definition for $\mathscr{P}$ and from Eq. (1.2) it follows that

$$\mathscr{P}K = \sum_{\gamma} |\gamma\rangle\langle\gamma|K|\gamma\rangle\langle\gamma| \text{ or } \mathscr{P} = \sum_{\gamma} |\gamma\rangle\langle\gamma|D_\gamma. \tag{2.2}$$

By $1-\mathscr{P}$ we mean everywhere $\mathscr{I}-\mathscr{P}$, where $\mathscr{I}$ is the identity operator. Expressions like $\mathscr{P}K$ and $(1-\mathscr{P})K$ indicate that $\mathscr{P}$ and $1-\mathscr{P}$ are *superoperators* which act on the operators K. As indicated by Fano,[7] this description can be formalized by picturing all operators K as elements of another Hilbert space, called the *Liouville space*. Superoperators will then represent transformations of K in the Liouville space. All superoperators will be represented by script letters, except Λ. Besides the superoperators $\mathscr{P}$ and $\mathscr{I}$, we will also meet the superoperators $\mathscr{L}$ (Liouville operator corresponding to H), $\mathscr{L}^0$ (Liouville operator corresponding to H^0), and the master operator Λ_d (Sec. 1). All superoperators have a tetradic representation (see Appendix A), due to the fact that the Liouville space is isomorphic with the space $\mathscr{H}\otimes\overline{\mathscr{H}}$, where $\overline{\mathscr{H}}$ is the dual of $\mathscr{H}$ and where $\otimes$ denotes the direct product. In Part I we also introduced the diagonal Liouville space, being the subspace containing only diagonal operators K_d. Clearly the superoperator $\mathscr{P}$ is the identity operator $\hat{I}$ of diagonal Liouville space as is seen by comparing the second expression of Eq. (2.2) with Ref. 1, Eq. (6.7).

The projection operators $\mathscr{P}$ and $1-\mathscr{P}$ satisfy the following properties:

$$\mathscr{P}^2 = \mathscr{P}, \quad \mathscr{P}(1-\mathscr{P}) = 0, \quad (1-\mathscr{P})^2 = 1-\mathscr{P}. \tag{2.3}$$

For the proofs, consider

$$\mathscr{P}^2K = \sum_{\gamma} |\gamma\rangle\langle\gamma|\{\sum_{\gamma'} |\gamma'\rangle\langle\gamma'|K|\gamma'\rangle\langle\gamma'|\}|\gamma\rangle\langle\gamma| = \sum_{\gamma\gamma'} |\gamma\rangle\delta_{\gamma\gamma'}\langle\gamma'|K|\gamma'\rangle\langle\gamma| = \mathscr{P}K. \tag{2.4}$$

The other properties of Eq. (2.3) follow from the above. From the definition of $\mathscr{P}K$ as the diagonal part it is clear that the commutator with H^0 vanishes:

$$[\mathscr{P}K,H^0]=0, \text{ or } (\mathscr{P}K)H^0=H^0\mathscr{P}K, \tag{2.5}$$

Though our developments in this paper *will not require coarse graining*, we will nevertheless make some comments on coarse graining in order to vindicate the method of Part I and to make the connection. To that end we show that slightly coarse grained macroscopic operators have vanishing nondiagonal part for large times or small enough frequencies, i.e., in the classical frequency limit (1.5). In general, such coarse grained operators defy the proper observance of quantum effects, for which the nondiagonal parts are essential (see Sec. 7). Let $Z(\gamma)$ be the density of states for systems in which γ can be considered to be a near continuous set of quantum parameters that label the states. Employing the energy as one of the labels, we also write $|\gamma\rangle = |\alpha\epsilon\rangle$. Thus,

$$\begin{aligned}\mathscr{P}K &= \sum_\gamma |\gamma\rangle\langle\gamma|K|\gamma\rangle\langle\gamma| \\ &= \int d\gamma\, Z(\gamma)|\gamma\rangle\langle\gamma|K|\gamma\rangle\langle\gamma| \\ &= \int d\epsilon \int d\alpha\, Z(\alpha,\epsilon)|\alpha\epsilon\rangle\langle\alpha\epsilon|K|\alpha\epsilon\rangle\langle\alpha\epsilon|. \end{aligned}\tag{2.6}$$

If the energy is coarse grained, i.e., measured in cells $\Delta\epsilon_i$ (being an energy spread which has macroscopically no significance), we need the number of states $Z(\alpha,\epsilon_i) = \int_{\Delta\epsilon_i} Z(\alpha,\epsilon)d\epsilon$. We then obtain the "fuzzy diagonal part" of K. The complementary fuzzy nondiagonal part will be made up of states of different α and (or) different energy cells. We may argue that for the majority of contributions the energies will be different. The time-dependent operator of the fuzzy nondiagonal part is then rapidly oscillating[8]

$$\begin{aligned}K_{nd}(t) &\approx e^{iH^0t/\hbar}\sum_{\alpha\alpha'}\sum_{i\neq j}|\alpha\epsilon_i\rangle\langle\alpha\epsilon_i|K|\alpha'\epsilon_j\rangle\langle\alpha'\epsilon_j|e^{-iH^0t/\hbar} \\ &= \sum_{\alpha\alpha'}\sum_{i\neq j} e^{i(\epsilon_i-\epsilon_j)t/\hbar}|\alpha\epsilon_i\rangle\langle\alpha\epsilon_i|K|\alpha'\epsilon_j\rangle\langle\alpha'\epsilon_j|. \end{aligned}\tag{2.7}$$

Since $(\epsilon_i-\epsilon_j)t/\hbar > 1$, the exponential represents rapid oscillations. In linear response theory we need the Laplace transformed operators for which

$$\begin{aligned}\int_0^\infty e^{-st}e^{i(\epsilon_i-\epsilon_j)t/\hbar}dt &= \frac{i}{(\epsilon_i-\epsilon_j)/\hbar+is} \\ &\approx \frac{i\hbar}{\epsilon_i-\epsilon_j}+\pi\hbar\delta(\epsilon_i-\epsilon_j), \quad |s| \text{ small.}\end{aligned}\tag{2.8}$$

For $\epsilon_i \neq \epsilon_j$ the delta contribution yields zero and the principal part is very small. For large t or small $|s|$ we can therefore often dismiss the nondiagonal contributions. An exception occurs when the diagonal part is zero. This we encounter in the quantum mechanical description of the Hall effect, on which we will report in a future paper on applications.

B. Some theorems for projector algebra

The Liouville operator for quantum systems is defined by

$$\mathscr{L}K = (1/\hbar)[H,K]. \tag{2.9}$$

Since $H = H^0 + \lambda V$, we write accordingly $\mathscr{L} = \mathscr{L}^0 + \lambda\mathscr{L}^1$. Thus, we also have

$$\mathscr{L}^0K = (1/\hbar)[H^0,K], \quad \mathscr{L}^1K = (1/\hbar)[V,K]. \tag{2.9a}$$

The following exponential identity is well known:

Theorem 1:

$$\begin{aligned}&e^{i\mathscr{L}t}K = e^{iHt/\hbar}Ke^{-iHt/\hbar}; \\ &e^{i\mathscr{L}^0t}K = e^{iH^0t/\hbar}Ke^{-iH^0t/\hbar}.\end{aligned}\tag{2.10}$$

The proof follows directly from integration of the Heisenberg equation of motion, $dK(t)/dt - i\mathscr{L}K(t) = 0$. The above statements are also valid for complex $t = t_1 + it_2$.

Theorem 2: We have the following operator identity involving diagonal parts:

$$e^{\sum_\gamma |\gamma\rangle\langle\gamma| f(\gamma)} = \sum_\gamma |\gamma\rangle\langle\gamma| e^{f(\gamma)}. \tag{2.11}$$

This property was used in Appendix B of Part I. There is a similar property for the exponential superoperator.

Theorem 3:

$$\begin{aligned}e^{-\Lambda_d t}K &= \left(\exp\left\{-\sum_{\gamma\gamma'}|\gamma\rangle\langle\gamma|[W_{\gamma'\gamma}D_{\gamma'}-W_{\gamma\gamma'}D_\gamma]t\right\}\right)K \\ &= \sum_\gamma |\gamma\rangle\langle\gamma|e^{-Mt}\langle\gamma|K|\gamma\rangle,\end{aligned}\tag{2.12}$$

where M is the master operator in the function space for functions $f(\gamma)$:

$$Mf(\gamma) \equiv -\sum_{\gamma''}[W_{\gamma''\gamma}f(\gamma'')-W_{\gamma\gamma''}f(\gamma)]. \tag{2.13}$$

The proof follows by series expansion, using also the closure property $\sum_\gamma|\gamma\rangle\langle\gamma| = 1$.

The following two theorems are necessary for the projection operator calculus of the sections which follow.

Theorem 4:

$$e^{-it(1-\mathscr{P})\mathscr{L}}(1-\mathscr{P})K = (1-\mathscr{P})e^{-it\mathscr{L}(1-\mathscr{P})}K. \tag{2.14}$$

It is to be born in mind that $\mathscr{P}$ or $1-\mathscr{P}$ always works on everything to its right, except when indicated contrary by the use of parentheses. The proof of Eq. (2.14) follows by series expansion:

$$\begin{aligned}&e^{-it(1-\mathscr{P})\mathscr{L}}(1-\mathscr{P})K \\ &\quad= [1-it(1-\mathscr{P})\mathscr{L}+\tfrac{1}{2}(it)^2(1-\mathscr{P})\mathscr{L}(1-\mathscr{P}) \\ &\qquad\times\mathscr{L}+\ldots](1-\mathscr{P})K \\ &\quad= [(1-\mathscr{P})-it(1-\mathscr{P})\mathscr{L}(1-\mathscr{P})+\tfrac{1}{2}(it)^2 \\ &\qquad\times(1-\mathscr{P})\mathscr{L}(1-\mathscr{P})\mathscr{L}(1-\mathscr{P})+\ldots]K \\ &\quad= (1-\mathscr{P})[1-it\mathscr{L}(1-\mathscr{P})+\ldots]K \\ &\quad= (1-\mathscr{P})e^{-it\mathscr{L}(1-\mathscr{P})}K.\end{aligned}\tag{2.15}$$

Theorem 5:

$$e^{-it(1-\mathscr{P})\mathscr{L}}K = [e^{-it\mathscr{L}^0}+O(\lambda)]K, \tag{2.16}$$

providing

$$\mathscr{P}[K,H^0] = 0. \tag{2.17}$$

The proof is simple, since

$$e^{-it(1-\mathscr{P})\mathscr{L}}K = e^{it(1-\mathscr{P})\mathscr{L}^0}K + O(\lambda). \tag{2.18}$$

However, $\mathscr{P}\mathscr{L}^0K = 0$ if Eq. (2.17) is satisfied.

At this moment it may surprise the reader that the condition (2.17) is necessary. Since $\mathscr{P}$ cannot work on H^0, which is diagonal already, it seems that Eq. (2.17) is identical to Eq. (2.5), so that the condition (2.17) is superfluous. The problem involves here the *validity of the Dirac notation*, which assumes that there is no boundary term arising from Green's theorem, which in turn implies that no states occur that go outside of the space $\mathscr{H}$ determined by H^0 and the boundary conditions. To appreciate this difficulty, we restate the Dirac matrix elements in terms of the standard algebraic notation; we have

$$\langle\phi\,|\{K\,|\psi\rangle\}\equiv(K\psi,\phi), \tag{2.19a}$$

and

$$\{\langle\phi\,|K\}\,|\psi\rangle\equiv(\psi,K^\dagger\phi). \tag{2.19b}$$

Now,

$$(K\psi,\phi)-(\psi,K^\dagger\phi)=Q_K(\psi,\phi), \tag{2.20}$$

where $Q_K(\psi,\phi)$ is the bilinear concomitant involving ϕ and ψ at the boundaries of the volume under consideration. The associativity of the Dirac bilinear form, i.e., the equality of the left hand sides of Eqs. (2.19a) and (2.19b), depends critically on the condition that $Q_K\equiv 0$. Only in this case can we equally well operate to the right (in state space) as to the left (in dual space). When this property of associativity is guaranteed we have

$$\begin{aligned}\langle\gamma|[K,H^0]|\gamma\rangle &= \langle\gamma|K\overrightarrow{H^0}|\gamma\rangle-\langle\gamma|\overleftarrow{H^0}K|\gamma\rangle\\ &=\langle\gamma|K|\gamma\rangle\epsilon_\gamma-\epsilon_\gamma\langle\gamma|K|\gamma\rangle=0,\end{aligned} \tag{2.21}$$

so that

$$\mathscr{P}[K,H^0]=\sum_\gamma|\gamma\rangle\,\langle\gamma|[K,H^0]|\gamma\rangle=0. \tag{2.22}$$

However, if it is not *a priori* clear that the bilinear concomitant vanishes, the following more general statement is proven in Appendix B:

Theorem 6:

$$\mathscr{P}[K,H^0]=-\sum_\gamma|\gamma\rangle\,\langle\gamma|Q_{H^0}(K\gamma,\gamma), \tag{2.23}$$

where Q_{H^0} is the bilinear concomitant associated with the operator H^0. Similarly, if $K_1\ldots K_n$ are arbitrary operators and if L_d is a diagonal operator,

$$\mathscr{P}[K_1\ldots K_n,L_d]=-\sum_\gamma|\gamma\rangle\,\langle\gamma|Q_{L_d}(K_1\ldots K_n\gamma,\gamma). \tag{2.24}$$

Concordant with these results we must in some cases modify the general rules for projector algebra. From the Dirac form for the diagonal part [Eq. (2.2)], one finds for a succession of operators

$$\mathscr{P}(K_1\ldots K_nL_d)=\mathscr{P}(K_1\ldots K_n)L_d=L_d\mathscr{P}(K_1\ldots K_n), \tag{2.25}$$

where in the matrix element $\langle\gamma|K_1\ldots K_nL_d|\gamma\rangle$ the operator L_d worked to the right. Similarly, one obtains

$$\mathscr{P}(L_dK_1\ldots K_n)=L_d\mathscr{P}(K_1\ldots K_n), \tag{2.26}$$

where in the matrix element $\langle\gamma|L_dK_1\ldots K_n|\gamma\rangle$ L_d operated to the left. This operation is, however, not equivalent to the previous operation if $Q_{L_d}\not\equiv 0$. It can then be shown that Eq. (2.25) remains valid, but that (2.26) contains an extra term:

$$\mathscr{P}(L_dK_1\ldots K_n)=L_d\mathscr{P}(K_1\ldots K_n)+\sum_\gamma|\gamma\rangle\langle\gamma|Q_{L_d}(K_1\ldots K_n\gamma,\gamma). \tag{2.26'}$$

One notes that Eqs. (2.25) and (2.26') result in Eq. (2.24).

In view of these difficulties, two options are open. Either we can do the entire computation in standard algebraic notation, which is not appealing since no simple form for the projectors $|\gamma\rangle\,\langle\gamma|$ exists in that case, or we can clearly delineate under which circumstances the Dirac projector algebra involving the usual results (2.22), (2.25), and (2.26) is valid. These circumstances are easily stated. Since operators like H^0 and L_d are fully self-adjoint (not just Hermitian), the boundary conditions on ψ and the dual boundary conditions on ϕ, necessary for the bilinear concomitant $Q(\psi,\phi)$ to vanish, must be identical. Thus, $Q(K_1\ldots K_n\gamma,\gamma)=0$ if $K_1\ldots K_n\gamma$ satisfies periodic boundary conditions. We conclude that the *standard Dirac projector algebra is valid whenever the operators $K_1\ldots K_n$ are invariant against translation over the dimensions of the system*. This will turn out to be the case in most instances, though exceptions will be met.

C. Adjointness of superoperators

In Part I, Sec. 8.2, we defined the scalar product in Liouville space by $\{A,B\}=\mathrm{Tr}AB^\dagger$. Generally, any superoperator acting in Liouville space can be written as

$$S=\sum_i P_i\overrightarrow{\ }\overleftarrow{\ }Q_i, \tag{2.27}$$

where the first arrow indicates that P_i works to the right, while the second arrow indicates that Q_i works to the left (see also Appendix A). The adjoint is defined by

$$\{A,\mathscr{S}^\dagger B\}=\{\mathscr{S}A,B\}. \tag{2.28}$$

This is satisfied if

$$\mathscr{S}^\dagger=\sum_i P_i^\dagger\overrightarrow{\ }\overleftarrow{\ }Q_i^\dagger. \tag{2.29}$$

The proof is simple:

$$\{A,\mathscr{S}^\dagger B\}=\sum_i\mathrm{Tr}A\,(P_i^\dagger BQ_i^\dagger)^\dagger=\mathrm{Tr}AQ_iB^\dagger P_i \tag{2.30}$$

and

$$\{\mathscr{S}A,B\}=\sum_i\mathrm{Tr}P_iAQ_iB^\dagger\overset{\text{cycl.perm.}}{=}\sum_i\mathrm{Tr}AQ_iB^\dagger P_i. \tag{2.31}$$

The superoperators $\mathscr{I}$ (unit operator) and $\mathscr{P}$ (projection operator) are self-adjoint, the proof of which is trivial. Likewise, the Liouville operator is self-adjoint:

$$\begin{aligned}\mathscr{L}^\dagger&=(1/\hbar)[H^\dagger\overrightarrow{\ }\overleftarrow{\ }I-I\overrightarrow{\ }\overleftarrow{\ }H^\dagger]\\ &=(1/\hbar)[H\overrightarrow{\ }\overleftarrow{\ }I-I\overrightarrow{\ }\overleftarrow{\ }H]=\mathscr{L}.\end{aligned} \tag{2.32}$$

Also, the master operator Λ_d is self-adjoint:

$$\begin{aligned}\Lambda_d^\dagger&=-\sum_{\gamma\gamma'}[W_{\gamma'\gamma}|\gamma''\rangle\,\langle\gamma|\overrightarrow{\ }\overleftarrow{\ }|\gamma\rangle\,\langle\gamma''|\\ &\quad-W_{\gamma\gamma'}|\gamma\rangle\,\langle\gamma|\overrightarrow{\ }\overleftarrow{\ }|\gamma\rangle\,\langle\gamma|]\\ &=-\sum_{\gamma\gamma'}[W_{\gamma\gamma'}|\gamma\rangle\,\langle\gamma''|\overrightarrow{\ }\overleftarrow{\ }|\gamma''\rangle\,\langle\gamma|\end{aligned}$$

$$- W_{\gamma\gamma'}|\gamma\rangle\langle\gamma|^{\rightarrow\leftarrow}|\gamma\rangle\langle\gamma|]$$
$$= -\sum_{\gamma\gamma'}[W_{\gamma'\gamma}|\gamma\rangle\langle\gamma''|^{\rightarrow\leftarrow}|\gamma''\rangle\langle\gamma|$$
$$- W_{\gamma\gamma'}|\gamma\rangle\langle\gamma|^{\rightarrow\leftarrow}|\gamma\rangle\langle\gamma|] = \Lambda_d, \tag{2.33}$$

where we changed the summation indices in the first sum and used microscopic reversibility $W_{\gamma\gamma'} = W_{\gamma'\gamma}$.

3. THE KINETIC EQUATIONS FOR THE DIAGONAL AND NONDIAGONAL PARTS OF THE DENSITY OPERATOR FOR SYSTEMS IN AN EXTERNAL FIELD

Let H be the system Hamiltonian and $-AF(t)$ the effect of an external field, where A is an operator in the system, such as the charge moment $e\Sigma\mathbf{r}_i$, and where $F(t)$ is a complex function which denotes an external field. We assume that the field is switched on at $t = 0$. The von Neumann equation then reads

$$\frac{\partial\rho}{\partial t} + i\mathcal{L}\rho = \frac{\partial\rho}{\partial t} + \left(\frac{i}{\hbar}\right)[H,\rho]$$
$$= \left(\frac{i}{\hbar}\right)u(t)F(t)[A,\rho], \tag{3.1}$$

where u is the unit step function. We split

$$\rho = \rho_d + \rho_{nd}. \tag{3.2}$$

Then, by application of the projection operators $\mathcal{P}$ and $(1-\mathcal{P})$ on Eq. (3.1), we obtain

$$\frac{\partial\rho_d}{\partial t} + i\mathcal{P}\mathcal{L}\rho_d + i\mathcal{P}\mathcal{L}\rho_{nd}$$
$$= (i/\hbar)F(t)\mathcal{P}[A,\rho_d] + (i/\hbar)F(t)\mathcal{P}[A,\rho_{nd}], \tag{3.3}$$

$$\frac{\partial\rho_{nd}}{\partial t} + i(1-\mathcal{P})\mathcal{L}\rho_d + i(1-\mathcal{P})\mathcal{L}\rho_{nd}$$
$$= (i/\hbar)F(t)(1-\mathcal{P})[A,\rho_d] + (i/\hbar)F(t)$$
$$\times(1-\mathcal{P})[A,\rho_{nd}]. \tag{3.4}$$

The following simplifications can be made: For the second term on the left of Eq. (3.3) we have

$$\mathcal{P}\mathcal{L}\rho_d = (1/\hbar)\mathcal{P}[H^0,\rho_d] + (\lambda/\hbar)\mathcal{P}[V,\rho_d] = 0, \tag{3.5}$$

That $[H^0,\rho_d] = 0$ is obvious. For the second term we have $\mathcal{P}[V,\rho_d] = [\mathcal{P}V,\rho_d]$ since the bilinear concomitant $Q_{\rho_d}(V\gamma,\gamma)$ vanishes, the potential V being invariant against translation of all q_i [typically, V is a two-body interaction potential $\Sigma_{i>j}v(\mathbf{r}_i - \mathbf{r}_j)$ (see Sec. 8 for examples)]. Further, we set $\mathcal{P}V = 0$; if V would have a diagonal part, we incorporate it in H^0, *which is to be chosen as the largest sub-Hamiltonian that can be diagonalized.*[9] In accord with the above, the second term on the left of Eq. (3.4) becomes simply $i\mathcal{L}\rho_d$. Further, it was found, in view of the later linearization, that it is best not to split up ρ in the field dependent terms.[10] Thus, Eqs. (3.3) and (3.4) simplify to

$$\frac{\partial\rho_d}{\partial t} + i\mathcal{P}\mathcal{L}\rho_{nd} = \left(\frac{i}{\hbar}\right)F(t)\mathcal{P}[A,\rho], \tag{3.3'}$$

$$\frac{\partial\rho_{nd}}{\partial t} + i\mathcal{L}\rho_d + i(1-\mathcal{P})\mathcal{L}\rho_{nd} = \left(\frac{i}{\hbar}\right)F(t)(1-\mathcal{P})[A,\rho].$$

We first solve formally Eq. (3.4'). The Green's "function" (more precisely the Green's superoperator) for Eq. (3.4') is defined by

$$\frac{\partial\mathcal{G}(t,t')}{\partial t} + i(1-\mathcal{P})\mathcal{L}\mathcal{G}(t,t') = \mathcal{I}\delta(t-t'). \tag{3.6}$$

The solution is

$$\mathcal{G}(t,t') \equiv \mathcal{G}(t-t') = u(t-t')e^{-i(t-t')(1-\mathcal{P})\mathcal{L}}. \tag{3.7}$$

With this Green's function the solution of (3.4') is

$$\rho_{nd}(t) = \mathcal{G}(t,0)\rho_{nd}(0)$$
$$+ \int_0^t dt'\,\mathcal{G}(t-t')$$
$$\times\{-i\mathcal{L}\rho_d(t') + (i/\hbar)F(t')(1-\mathcal{P})[A,\rho(t')]\}. \tag{3.8}$$

This result is now substituted into (3.3'); we then arrive at the integro-differential operator equation

$$\frac{\partial\rho_d}{\partial t} + \mathcal{P}\mathcal{L}\int_0^t dt'\,\mathcal{G}(t-t')\mathcal{L}\rho_d(t')$$
$$= \left(\frac{i}{\hbar}\right)F(t)\mathcal{P}[A,\rho(t)] + \left(\frac{1}{\hbar}\right)\mathcal{P}\mathcal{L}$$
$$\times\int_0^t dt'F(t')\mathcal{G}(t-t')(1-\mathcal{P})[A,\rho(t')]$$
$$- i\mathcal{P}\mathcal{L}\mathcal{G}(t,0)\rho_{nd}(0). \tag{3.9}$$

Equations (3.8) and (3.9) give the full evolution of the diagonal and nondiagonal parts. *These equations are fully equivalent to the original von Neumann equation*, though the occurrence of $\rho_{nd}(0)$ indicates that a fixed direction in time has been introduced. Note that we do not make an initial random phase assumption. The equations are still non-Markovian in that they contain all memory effects as evidenced by the time convolution integral.

We now make the standard simplification of *linear* response theory; i.e., in the field terms of (3.8) and (3.9) we replace $\rho(t)$ by ρ_{eq}, where for the present, ρ_{eq} is the canonical (or grand canonical) distribution in the full dynamics,

$$\rho_{eq} = e^{-\beta H}/z, \quad z = \mathrm{Tr}e^{-\beta H}, \quad \beta = 1/kT. \tag{3.10}$$

Thus (3.8) and (3.9) become

$$\frac{\partial\rho_d}{\partial t} + \mathcal{P}\mathcal{L}\int_0^t dt'\,\mathcal{G}(t-t')\mathcal{L}\rho_d(t')$$
$$= \left(\frac{i}{\hbar}\right)F(t)\mathcal{P}[A,\rho_{eq}] + \left(\frac{1}{\hbar}\right)\mathcal{P}\mathcal{L}$$
$$\times\int_0^t dt'F(t')\mathcal{G}(t-t')(1-\mathcal{P})[A,\rho_{eq}]$$
$$- i\mathcal{P}\mathcal{L}\mathcal{G}(t,0)\rho_{nd}(0), \tag{3.11}$$

$$\rho_{nd}(t) = \int_0^t dt'\,\mathcal{G}(t-t')$$
$$\times\left\{-i\mathcal{L}\rho_d(t') + \left(\frac{i}{\hbar}\right)F(t')(1-\mathcal{P})[A,\rho_{eq}]\right\}$$
$$+ \mathcal{G}(t,0)\rho_{nd}(0). \tag{3.12}$$

In the next section we shall carry through the van Hove limit. This will have two results: The equations for ρ_d and ρ_{nd} become decoupled, and the equation for ρ_d will become Markovian.

4. THE VAN HOVE LIMIT

A. Kinetic equation for the diagonal part

The memory term [second term on the lhs of Eq. (3.11)] is basically treated as in Zwanzig's paper.[11] However, we will show the explicit emergence of the operator Λ_d for this term. We notice in advance that the only operators occurring in this term are H^0, ρ, and V. Of ρ and V we assume that they are invariant against translation over the dimensions of the system. Since ρ is a bounded operator,[12] we can let the system become very large, so that the assumption of translational invariance is no ultimate physical restriction. As to V, usually it is a two-body interaction potential so that it is translationally invariant; note that translation means a shift in the, say, x coordinates, of *all* the particles involved in H^0. Because of the above assumptions, the rules of Dirac projector algebra fully apply. To start with, we have

$$\mathscr{L}\rho_d(t') = \lambda\mathscr{L}^1\rho_d(t') = (\lambda/\hbar)[V,\rho_d(t')], \tag{4.1}$$

since $\mathscr{L}^0\rho_d = 0$. Next we use Theorem 5; thus, with Eq. (3.7) we find

$$\mathscr{G}(t-t')\mathscr{L}\rho_d(t') = \lambda\,[\mathscr{G}_0(t-t') + O(\lambda)]\mathscr{L}^1\rho_d(t'), \tag{4.2}$$

where

$$\mathscr{G}_0(t) = \mathscr{J}u(t)e^{-i\mathscr{L}^0 t}. \tag{4.3}$$

As indicated above, condition (2.17) is fulfilled. We must now operate on Eq. (4.2) with $\mathscr{P}\mathscr{L} = \mathscr{P}\mathscr{L}^0 + \lambda\mathscr{P}\mathscr{L}^1$; we have

$$\mathscr{P}\mathscr{L}^0[\text{rhs of Eq. (4.2)}] = (\lambda/\hbar)\mathscr{P}[H^0,\mathscr{G}_0(t-t')\mathscr{L}^1\rho_d(t')] = 0, \tag{4.4}$$

since $\mathscr{P}\mathscr{L}^1\rho_d(t') = 0$. We are thus left with

$$\lambda\mathscr{P}\mathscr{L}^1[\text{rhs of Eq. (4.2)}] = (\lambda^2/\hbar^2)\mathscr{P}[V,\mathscr{G}_0(t-t')[V,\rho_d(t')]]. \tag{4.5}$$

We now employ the explicit form for $\mathscr{P}$ [Eq. (2.2)], we substitute for $\mathscr{G}_0$ from Eq. (4.3), and we use Eq. (2.10); we thus arrive at

$$\mathscr{P}\mathscr{L}\int_0^t dt'\,\mathscr{G}(t-t')\mathscr{L}\rho_d(t') = \sum_\gamma |\gamma\rangle\langle\gamma|(\lambda^2/\hbar^2)\int_0^t dt'\langle\gamma|[V,e^{-iH^0(t-t')/\hbar}[V,\rho_d(t')]e^{iH^0(t-t')/\hbar}]|\gamma\rangle. \tag{4.6}$$

The matrix element is, with $p(\gamma,t) = \langle\gamma|\rho_d(t)|\gamma\rangle$,

$$\begin{aligned}\langle\gamma|[V,\ldots]|\gamma\rangle = \sum_{\gamma''}\{&\langle\gamma|V|\gamma''\rangle e^{-i\epsilon_{\gamma'}(t-t')/\hbar}[\langle\gamma''|V|\gamma\rangle p(\gamma,t') - p(\gamma'',t')\langle\gamma''|V|\gamma\rangle]e^{i\epsilon_\gamma(t-t')/\hbar} - e^{-i\epsilon_\gamma(t-t')/\hbar}\\ &\times[\langle\gamma|V|\gamma''\rangle p(\gamma'',t') - p(\gamma,t')\langle\gamma|V|\gamma''\rangle]e^{i\epsilon_{\gamma''}(t-t')/\hbar}\langle\gamma''|V|\gamma\rangle\}\\ = 2\sum_{\gamma''}&|\langle\gamma|V|\gamma''\rangle|^2\cos[(\epsilon_{\gamma''}-\epsilon_\gamma)(t-t')/\hbar][p(\gamma,t') - p(\gamma'',t')].\end{aligned} \tag{4.7}$$

Thus, the memory term becomes from Eq. (4.6) and (4.7), denoting it as $(\partial\rho_d/\partial t)_{\text{mem.}}$,

$$\left(\frac{\partial\rho_d}{\partial t}\right)_{\text{mem.}} = \sum_{\gamma\gamma''}|\gamma\rangle\langle\gamma|\frac{2\lambda^2}{\hbar^2}|\langle\gamma|V|\gamma''\rangle|^2\int_0^t dt'\cos[(\epsilon_{\gamma''}-\epsilon_\gamma)(t-t')/\hbar][p(\gamma,t') - p(\gamma'',t')], \tag{4.8}$$

with Laplace transform

$$\Psi(s) = \sum_{\gamma\gamma''}|\gamma\rangle\langle\gamma|\frac{2\lambda^2}{\hbar^2}|\langle\gamma|V|\gamma''\rangle|^2\frac{s}{s^2+(\epsilon_{\gamma''}-\epsilon_\gamma)^2/\hbar^2}[P(\gamma,s) - P(\gamma'',s)]. \tag{4.9}$$

Now, for small $|s|$, i.e., large t in the van Hove limit,

$$\frac{s}{s^2+(\epsilon_{\gamma''}-\epsilon_\gamma)^2/\hbar^2} = \frac{1}{2i}\left[\frac{1}{(\epsilon_{\gamma''}-\epsilon_\gamma)/\hbar - is} - \frac{1}{(\epsilon_{\gamma''}-\epsilon_\gamma)/\hbar + is}\right] \approx \pi\hbar\delta(\epsilon_{\gamma''}-\epsilon_\gamma). \tag{4.10}$$

Hence,

$$\Psi(s) \approx \sum_{\gamma\gamma''}|\gamma\rangle\langle\gamma|[W_{\gamma\gamma''}P(\gamma,s) - W_{\gamma''\gamma}P(\gamma'',s)], \tag{4.11}$$

where we used Eq. (1.3). Transforming back, we obtain for the memory term

$$\begin{aligned}\left(\frac{\partial\rho_d}{\partial t}\right)_{\text{mem.}} &= \sum_{\gamma\gamma''}|\gamma\rangle\langle\gamma|[W_{\gamma\gamma''}p(\gamma,t) - W_{\gamma''\gamma}p(\gamma'',t)]\\ &= \sum_{\gamma\gamma''}|\gamma\rangle\langle\gamma|[W_{\gamma\gamma''}D_\gamma - W_{\gamma''\gamma}D_{\gamma''}]\rho_d\\ &= \Lambda_d\rho_d,\end{aligned} \tag{4.12}$$

with Λ_d being the master operator (1.1). We noted before that Λ_d gives the effect of the interaction λV which determines the transition probabilities $W_{\gamma\gamma''}$ of this operator. *So in the van Hove limit the memory term leads to the occurrence of relaxation due to the interaction λV.*

We now turn to the two terms on the rhs of Eq. (3.11). Since $A = A(\{q\})$ is generally not invariant against a translation of the system, we must be careful with the projector operations; in particular we note that $\mathscr{P}[A,\rho_{\text{eq}}] \neq 0$. We will make two assumptions on A:

(i) A commutes with the interaction potential λV. This is the case for all conductivity problems for which $A = e\Sigma\mathbf{r}_i$ is a position operator. For spin relaxation phenomena the assumption entails that we neglect the commutator with the nonsecular part of the spin–spin interaction Hamiltonian.

(ii) $\dot{A}$ is translational invariant. Again, this is certainly true for conductivity problems where $\dot{A}$ represents the total momentum; in magnetic problems in periodic structures this is likewise fulfilled.

For ρ_{eq} we have the limit

$$\rho_{eq} = \lim_{\lambda \to 0} e^{-\beta H}/z = e^{-\beta H^0}/z^0, \quad z^0 = \mathrm{Tr} e^{-\beta H^0}. \tag{4.13}$$

The first term on the right-hand side of (3.11) is calculated from Kubo's lemma

$$\begin{aligned}[A, e^{-\beta H}] &= -\int_0^\beta d\beta'\, e^{-\beta H}[A^H(-i\hbar\beta'), H] \\ &= -i\hbar \int_0^\beta d\beta'\, e^{-\beta H}\dot{A}^H(-i\hbar\beta').\end{aligned} \tag{4.14}$$

The limit $\lambda \to 0$ yields for $[A, \exp(-\beta H)]$:

$$\begin{aligned}\mathscr{P}[A, e^{-\beta H^0}] &= -i\hbar \int_0^\beta d\beta'\, \mathscr{P} e^{-\beta H^0} e^{\hbar\beta' \mathscr{L}^0} \dot{A}^I(0) \\ &= -i\hbar \int_0^\beta d\beta'\, e^{-\beta H^0} e^{\hbar\beta' \mathscr{L}^0} (\dot{A}^I(0))_d,\end{aligned} \tag{4.15}$$

where $\dot{A}^I(t)$ is the interaction operator, see (2.10),

$$\dot{A}^I(t) = e^{i\mathscr{L}^0 t} \dot{A} = e^{iH^0 t/\hbar} \dot{A} e^{-iH^0 t/\hbar}. \tag{4.16}$$

The $\mathscr{P}$ operator could be carried through the diagonal operators in the second member of (4.15) because of assumption (ii) above. Since $\exp(\hbar\beta' \mathscr{L}^0)(\dot{A})_d = (\dot{A})_d$ (as seen by series expansion), the integration over $d\beta'$ can be carried out. So the relevant term of (3.11) becomes

$$(i/\hbar) F(t) \mathscr{P}[A, \rho_{eq}] = \beta F(t) \rho_{eq} (\dot{A})_d. \tag{4.17}$$

[Note we write $(\dot{A})_d$ to indicate that the differentiation comes prior to the diagonal part projection; whereas $(\dot{A})_d \neq 0$, we have $\dot{A}_d = (d/dt) A_d = 0$ due to the Heisenberg equation of motion.]

Next we consider the second term on the rhs of (3.11). This term represents the effect of relaxation due to collisions on the field-caused motion (streaming). For ρ_{eq} we cannot immediately use the limiting form (4.13) since this term involves a time integral, so that the van Hove limit must be carried out judiciously. The commutator of A with ρ_{eq} is again written with the aid of Kubo's lemma (4.14), i.e., as an integral over $\dot{A}$. The $\mathscr{P}\mathscr{L}$ in front of the term is split into $\mathscr{P}\mathscr{L}^0 + \lambda \mathscr{P}\mathscr{L}^1$. For $\mathscr{G}$ we write $\mathscr{G}^0 + 0(\lambda)$ with $\mathscr{G}^0$ being diagonal, see (4.3). Because of assumption (ii), the $\mathscr{P}$ of $\mathscr{P}\mathscr{L}^0$ can now be carried through all factors till it meets $1 - \mathscr{P}$ which it annihilates. Thus the $\mathscr{P}\mathscr{L}^0$ operation yields zero. We therefore arrive at

second field term

$$\begin{aligned}&= -(1/\hbar)\lambda \mathscr{P}\mathscr{L}^1 \int_0^t dt' F(t')[e^{-i\mathscr{L}^0(t-t')} + \mathscr{O}(\lambda)] \\ &\quad \times (1 - \mathscr{P}) \int_0^\beta d\beta' e^{-\beta H}[A^H(-i\hbar\beta'), H]/z.\end{aligned} \tag{4.18}$$

The limits $e^{-\beta H} \to e^{-\beta H^0}$, $z \to z^0$ can safely be taken, so $(1 - \mathscr{P})$ is moved in front of the commutator. In accord with the rules developed in Appendix B, we must evaluate the commutator prior to nondiagonal projection. We thus write

$$\begin{aligned}&(1 - \mathscr{P})[A^H(-i\hbar\beta'), H] \\ &\quad = \hbar i (1 - \mathscr{P}) \dot{A}^H(-i\hbar\beta') \\ &\quad = \hbar i \dot{A}^H(-i\hbar\beta') - \hbar i (\dot{A}^H(-i\hbar\beta'))_d \\ &\quad = [A^H(-i\hbar\beta'), H] - [A_d^H(-i\hbar\beta'), \lambda V].\end{aligned} \tag{4.19}$$

Now for classical frequencies $\hbar\omega \ll kT$ the quantum correction factor due to the argument $-i\hbar\beta'$ does not apply, see Paper I, Sec. 9. Thus in (4.19) we make the approximation

$$\begin{aligned}[A^H(-i\hbar\beta'), H] &\approx [A^H(0), H] \\ &= [A, H^0] + [A, \lambda V] = [A, H^0],\end{aligned} \tag{4.20}$$

where we used assumption (i). Thus,

$$\begin{aligned}&(1 - \mathscr{P})[A^H(-i\hbar\beta'), H] \\ &\quad \approx [A, H^0] - [A_d^H(-i\hbar\beta'), \lambda V] \\ &\quad \approx [A, H^0] - [A_d^I(-i\hbar\beta'), \lambda V],\end{aligned} \tag{4.21}$$

the last equality being justified, of course, for small λ. We further note $A_d^I(-i\hbar\beta') = e^{\mathscr{L}^0 \hbar\beta'} A_d = A_d$. Substituting the last term of (4.21), labelled (b), into (4.18), we find

second field term (b)

$$= \frac{\lambda^2 \beta}{\hbar z^0} \mathscr{P}[V, \int_0^t dt' F(t') e^{-i\mathscr{L}^0(t-t')} e^{-\beta H^0}[A_d, V]] \tag{4.22}$$

where terms $\mathscr{O}(\lambda^3)$ are ignored. We now use the explicit form for $\mathscr{P}$ of (2.2). We then obtain

second term (b)

$$\begin{aligned}&= \frac{\lambda^2 \beta}{\hbar^2 z^0} \sum_{\gamma\gamma''} |\gamma\rangle\langle\gamma| \Big\{ \langle\gamma|V|\gamma''\rangle \int_0^t dt' F(t') \exp[-i\epsilon_{\gamma''}(t-t')/\hbar] \exp(-\beta\epsilon_{\gamma''}) \\ &\quad \times [\langle\gamma''|A_d|\gamma''\rangle \langle\gamma''|V|\gamma\rangle - \langle\gamma''|V|\gamma\rangle \langle\gamma|A_d|\gamma\rangle] e^{i\epsilon_\gamma(t-t')/\hbar} \\ &\quad - \int_0^t dt' F(t') e^{-i\epsilon_\gamma(t-t')/\hbar} e^{-\beta\epsilon_\gamma} [\langle\gamma|A_d|\gamma\rangle \langle\gamma|V|\gamma''\rangle - \langle\gamma|V|\gamma''\rangle \langle\gamma''|A_d|\gamma''\rangle] \exp[i\epsilon_{\gamma''}(t-t')/\hbar] \langle\gamma''|V|\gamma\rangle \Big\} \\ &= \frac{2\lambda^2 \beta}{\hbar^2 z^0} \sum_{\gamma\gamma''} |\gamma\rangle\langle\gamma| |\langle\gamma|V|\gamma''\rangle|^2 \int_0^t dt' F(t') \cos[(\epsilon_{\gamma''} - \epsilon_\gamma)(t-t')/\hbar] \\ &\quad \times [\langle\gamma''|A_d|\gamma''\rangle \exp(-\beta\epsilon_{\gamma''}) - \langle\gamma|A_d|\gamma\rangle \exp(-\beta\epsilon_\gamma)].\end{aligned} \tag{4.23}$$

The Laplace transform is

$$\chi(s) = \frac{2\lambda^2 \beta}{\hbar^2 z^0} \sum_{\gamma\gamma''} |\gamma\rangle\langle\gamma| |\langle\gamma|V|\gamma''\rangle|^2 F(s) \frac{s}{s^2 + (\epsilon_{\gamma''} - \epsilon_\gamma)^2/\hbar^2} [\langle\gamma''|A_d|\gamma''\rangle \exp(-\beta\epsilon_{\gamma''}) - \langle\gamma|A_d|\gamma\rangle \exp(-\beta\epsilon_\gamma)]. \tag{4.24}$$

With the limit expression (4.10) we obtain upon transforming back, using also (1.3),

second term (b)

$$= \beta F(t) \sum_{\gamma\gamma''} |\gamma\rangle \langle\gamma| W_{\gamma\gamma''} [\langle\gamma''|A_d|\gamma''\rangle - \langle\gamma|A_d|\gamma\rangle] e^{-\beta\epsilon_\gamma}/z^0$$
$$= -\beta F(t) \rho_{eq} \Lambda_d A_d. \tag{4.25}$$

We can now do the same for part (a), the first term on the right-hand side of (4.21). We obtain,

second field term (a)

$$= \beta\lambda i z^0 \int_0^t dt' F(t') [V, e^{-\beta H^0} \dot{A}^I_{nd}(t'-t)]. \tag{4.26}$$

The term is of order $\lambda \dot{A}_{nd}$. It is a nondiagonal contribution which can be dismissed in the diagonal result. In the full theory it is still negligible, being of order λ, compared to the nondiagonal field term (4.34).

Finally, there is the third term on the right-hand side of (3.11). $\mathscr{L}$ and $\mathscr{G}$ are written as $\mathscr{L}^0 + \mathscr{O}(\lambda)$ and $\mathscr{G}^0 + \mathscr{O}(\lambda)$. Since ρ is assumed to be invariant to translation over the dimensions of the system, $\mathscr{P}$ can be moved to the right till is annihilates $\rho_{nd}(0)$. Thus this term does not contribute. We conclude that for the equation for the diagonal part *an initial random phase assumption is not necessary.*

Both field terms, (4.17) and (4.25), are now combined to read

field terms

$$= \beta F(t) \rho_{eq} (-\Lambda_d A_d + (\dot{A})_d) = \beta F(t) \rho_{eq} (\dot{A}^R)_d, \tag{4.27}$$

where we introduced the reduced derivative operator or current operator,[13]

$$J_{A,d} \equiv (\dot{A}^R)_d = -\Lambda_d A_d + (\dot{A})_d. \tag{4.28}$$

The reduced derivative operator is, of course, defined as

$$\dot{A}^R = \left[\lim_{\lambda,t} \dot{A}^H(t) \right]_{t \to 0+}, \tag{4.29}$$

where $\lim_{\lambda,t}$ denotes the van Hove limit. The result (4.28) follows from *I* Eq. (6.37).[14] The quantity J_A is the macroscopic current associated with the transport $\dot{A}$. The two parts will be labelled *pondermotive current* [second term of (4.28)] and *collisional current* [first term of (4.28)]. This latter type of current has not yet generally been recognized as existing. More will be said on this in Sec. 8 [see in particular Eq. (8.47a)]. {We still note that the second contribution of (4.28) is due to A not being translationally invariant. For operators B which are translationally invariant, this term does not occur, for then $(\dot{B})_d = (1/\hbar i)\mathscr{P}[B, H^0]$ $= (1/\hbar i)[\mathscr{P}B, H^0] = 0$. Thus (4.28) is not incontradiction with (6.6).}

We now collect the various results, (4.12) and (4.27). For sufficiently small fields F, the diagonal equation is found to be

$$\frac{\partial \rho_d^R(t)}{\partial t} + \Lambda_d \rho_d^R(t) = \beta F(t) \rho_{eq} [-\Lambda_d A_d + (\dot{A})_d]. \tag{4.30}$$

We can also write this in terms of $p^R(\gamma,t) = \langle\gamma|\rho_d^R(t)|\gamma\rangle$; with the explicit expressions for Λ_d Eq. (1.1), we find

$$\frac{\partial p^R(\gamma,t)}{\partial t} + Mp^R(\gamma,t) = \beta F(t) \rho_{eq}(\gamma) \langle\gamma|(\dot{A}^R)_d|\gamma\rangle, \tag{4.31}$$

where

$$Mp^R(\gamma,t) = \sum_{\gamma''} [W_{\gamma\gamma''} p^R(\gamma,t) - W_{\gamma''\gamma} p^R(\gamma'',t)]. \tag{4.32}$$

The left-hand side of (4.31) is as in Zwansig's paper.[11] The streaming terms are new. The operator form (4.30) is more lucid for the application of linear response theory.

The formal solution is with the initial density operator $\rho(0) = \rho_{eq}$,

$$\rho_d^R(t) = \rho_{eq} + \beta u(t) \int_0^t dt' F(t') e^{-\Lambda_d(t-t')} \rho_{eq} (\dot{A}^R)_d. \tag{4.33}$$

The second part will be denoted as $\Delta\rho_d(t)$.

Equation (4.30) or Eq. (4.31) will be referred to as the *inhomogeneous master equation*. These equations give the effect of the interaction λV as evidenced by the relaxation term, as well as of the external field. In the absence of the field it reduces to the ordinary (homogeneous) master equation. This equation is Markovian and exhibits irreversibility, on which we comment in Sec. 5.

B. Kinetic equation for the nondiagonal part

The first term of (3.12) is easily seen to be of order λ so it will be dismissed. For the second term of (3.12) we use again Kubo's lemma. So we obtain

$$(i/\hbar) \int_0^t dt' F(t') \mathscr{G}(t-t')(1-\mathscr{P})[A, \rho_{eq}]$$
$$= \int_0^t dt' F(t') e^{-i\mathscr{L}^0(t-t')}$$
$$\times \int_0^\beta d\beta' \rho_{eq} e^{\mathscr{L}^0 \hbar\beta'} (\dot{A})_{nd} + \mathscr{O}(\lambda). \tag{4.34}$$

The reduced nondiagonal operators show no relaxation and we have

$$J_{A,nd} \equiv (\dot{A}^R)_{nd} = (\dot{A})_{nd} \tag{4.35}$$

the extreme rhs being the Schrödinger operator.[8] The final nondiagonal result thus becomes

$$\rho_{nd}^R(t) = \int_0^t dt' F(t') \int_0^\beta d\beta' e^{-i\mathscr{L}^0(t-t')} \rho_{eq}$$
$$\times e^{\mathscr{L}^0 \hbar\beta'} (\dot{A}^R)_{nd} + e^{-i\mathscr{L}^0 t} \rho_{nd}(0). \tag{4.36}$$

Differentiating, we obtain

$$\frac{\partial \rho_{nd}^R(t)}{\partial t} + i\mathscr{L}^0 \rho_{nd}^R(t) = F(t) \rho_{eq} \int_0^\beta d\beta' e^{\mathscr{L}^0 \hbar\beta'} (\dot{A}^R)_{nd}. \tag{4.37}$$

We find that $\rho_{nd}(0)$ drops out again. We note that now ρ^R_{nd} became decoupled of ρ^R_d. Thus, in the van Hove limit ρ^R_d and ρ^R_{nd} evolve independently. While the behavior for ρ^R_d became Markovian, this is not so for the behavior of ρ^R_{nd}. On the contrary, Eq. (4.37) has the structure of the original von Neumann equation; in particular, we note that it is time-reversible. We will refer to Eq. (4.37) as the *inhomogeneous interaction equation*. In the absence of an external field, Eq. (4.37) will be called the (homogeneous) interaction equation. We further note that we did not use the asymptotic time property of the van Hove limit. Indeed, there is no necessity for this in the nondiagonal part.

C. Kinetic equation for ρ^R_{total} in the van Hove Limit

Lastly, we find a kinetic equation for the total ρ^R. To this effect we first note that Λ_d destroys any nondiagonal operator[4],

$$\Lambda_d(1-\mathscr{P})K=0. \tag{4.38}$$

In particular, $\Lambda_d\rho^R_{nd}=0$. Thus, Eq. (4.30) also reads

$$\frac{\partial\rho^R_d(t)}{\partial t}+\Lambda_d\rho^R(t) = F(t)\rho_{\text{eq}}\int_0^\beta d\beta' e^{\hbar\beta'\mathscr{L}^0}(\dot{A})^R_d. \tag{4.39}$$

For (4.37) we also have since $\mathscr{L}^0\rho^R_d=0$,

$$\frac{\partial\rho^R_{nd}(t)}{\partial t}+i\mathscr{L}^0\rho^R(t) = F(t)\rho_{\text{eq}}\int_0^\beta d\beta' e^{\hbar\beta'\mathscr{L}^0}(\dot{A}^R)_{nd}. \tag{4.40}$$

Upon adding we obtain

$$\frac{\partial\rho^R(t)}{\partial t}+(\Lambda_d+i\mathscr{L}^0)\rho^R(t)=F(t)\rho_{\text{eq}}\int_0^\beta d\beta' e^{\hbar\beta'\mathscr{L}^0}\dot{A}^R. \tag{4.41}$$

The total kinetic operator is of the damped oscillator type; we thus established [compare Eq. (4.41) with (3.1)]

$$\lim_{\lambda,t} i\mathscr{L}=\Lambda_d+i\mathscr{L}^0\equiv\Lambda. \tag{4.42}$$

More correctly, the time dependence must be brought in, i.e.,

$$\lim_{\lambda,t}\exp(\pm i\mathscr{L}t)=\exp[(-\Lambda_d\pm i\mathscr{L}^0)t] =\exp(-\Lambda^{(\dagger)}t), \tag{4.43}$$

or in terms of the resolvent

$$\lim_{\lambda,s}\frac{1}{\pm i\mathscr{L}-s}=\frac{1}{\pm i\mathscr{L}^0-\Lambda_d-s}=\frac{1}{-\Lambda^{(\dagger)}-s}. \tag{4.44}$$

Thus, the van Hove limit has drastically altered the time behavior of $\rho(t)$ and also of all Heisenberg operators $B^H(t)$, as we discuss in the next section. In Eq. (4.41) all references to diagonal and nondiagonal parts have disappeared. This separation can therefore be seen as a mere expediency to obtain the evolution operator Λ of the van Hove limit procedure. Equation (4.41) is the *complete inhomogeneous evolution equation*.

5. RESULTS IN THE ABSENCE OF AN EXTERNAL FIELD; IRREVERSIBILITY

A. Time dependence of $\rho^R_{sp}(t)$ and $B^R(t)$

We consider in this section the behavior in the absence of an external field, denoting the density operator for this case as ρ_{sp} (sp for spontaneous). Then for the original operator of the full von Neumann equation we have the well known result

$$\rho_{\text{sp}}(t)=e^{-iHt/\hbar}\rho(0)e^{iHt/\hbar}=e^{-i\mathscr{L}t}\rho(0), \tag{5.1}$$

while for the Heisenberg operators of the dynamical variables we have

$$B^H(t)=e^{iHt/\hbar}B^H(0)e^{-iHt/\hbar}=e^{i\mathscr{L}t}B, \tag{5.2}$$

where $B^H(0)=B^S$

Next we consider the solution of the complete evolution equation (4.41) for $F=0$. We find

$$\rho^R_{\text{sp}}(t)=e^{-(\Lambda_d+i\mathscr{L}^0)t}\rho(0), \tag{5.3}$$

To obtain the form for $B^R(t)$, where $B^R(t)$ is a reduced Heisenberg operator, we write the average $\langle B\rangle$ both in the Schrödinger and the Heisenberg picture; this gives

$$\text{Tr}[\rho(0)B^R(t)]=\text{Tr}[Be^{-(\Lambda_d+i\mathscr{L}^0)t}\rho(0)]. \tag{5.4}$$

We now use Lemma 5 of Appendix C; this carries Eq. (5.4) over into

$$\text{Tr}[\rho(0)B^R(t)]=\text{Tr}[\rho(0)e^{-(\Lambda_d-i\mathscr{L}^0)t}B], \tag{5.5}$$

which shows that[15]

$$B^R(t)=e^{-(\Lambda_d-i\mathscr{L}^0)t}B. \tag{5.6}$$

Though the behavior of Eq. (5.3) or (5.6) suggests that of a damped oscillator, we may not conclude that ρ^R_{sp} or B^R approaches zero for $t\to\infty$. In this respect the superoperators are treacherous. The point is that for any t we can write

$$e^{-(\Lambda_d\pm i\mathscr{L}^0)t}K=e^{-\Lambda_d t}K_d+e^{\mp i\mathscr{L}^0 t}K_{nd}, \tag{5.7}$$

as is easily found since $\exp(\mp i\mathscr{L}^0 t)K_d=K_d$ and $\exp(-\Lambda_d t)K_{nd}=K_{nd}$. To find the asymptotic limit of Eq. (5.7), we must make a spectral resolution of Λ_d and $\mathscr{L}^0$ (cf, Part I, Sec. 8). Λ_d has an isolated eigenvalue zero, which leads to the asymptotic (equilibrium) form of $\rho_{d,\text{sp}}$ and B^R_d; the nondiagonal part does not show convergence, except in the Riemann–Lebesque sense; compare I, Eqs. (4.3) and (4.4).

B. On the cause of irreversibility

We now discuss the cause of irreversibility. First we restate the well known fact that *there is no entropy production and thus no dissipation for the full microscopic system described by the total H*. The proof is simple. With $S=-k\,\text{Tr}(\rho\log\rho)$ we have from the von Neumann equation

$$\frac{dS}{dt}=-k\,\text{Tr}\left[(1+\log\rho)\frac{\partial\rho}{\partial t}\right] =-\frac{ki}{\hbar}\text{Tr}\{(1+\log\rho)[\rho,H]\}=0, \tag{5.8}$$

since by cyclic permutativity of the trace

$$\mathrm{Tr}[(1+\log\rho)\rho H] - \mathrm{Tr}[(1+\log\rho)H\rho]$$
$$= \mathrm{Tr}[\rho(\log\rho)H] - \mathrm{Tr}[(\log\rho)H\rho] = 0. \quad (5.9)$$

We must now carry out the determination of dS/dt for the ρ of the equations in this paper. First suppose we take ρ as given by Eqs. (3.8) and (3.9), with $F = 0$, prior to the van Hove limit. Clearly, if we use both equations, the entropy production is still zero, since these equations are fully equivalent to the original von Neumann equation. Suppose we now make an initial random phase assumption $\rho_{nd}(0) = 0$. This has no effect; for if, in the full evolution equations of Sec. 3, we set $\rho_{nd} = 0$ at $t = 0$, then at any $t_1 > 0$, $\rho_{nd}(t_1) \neq 0$; the time t_1 can be taken as new reference point, which nullifies the initial random phase assumption.

More remarkable, however, suppose we make a repeated random phase assumption $\rho_{nd}(t) \equiv 0$. Now the Green's integrals in both Eqs. (3.8) and (3.9) are zero, showing that $\partial\rho_d/\partial t = 0$. As a consequence, $dS/dt = 0$. Thus, contrary to common belief, the quantum mechanical developments of this paper show that a *repeated random phase assumption without further statements produces no irreversibility.*

When we now come to the results after the van Hove limit we find that there is entropy production associated with ρ_d; thus, for the "diagonal entropy"

$$dS_d/dt \geqslant 0. \quad (5.10)$$

The proof was given in Part I, Sec. 8.1. We thus proved the following:

(i) *A partitioning of the Hamiltonian $H = H^0 + \lambda V$, together with application of the van Hove limit and the large system limit, yields irreversibility, as evidenced by a positive definite entropy production*; (ii) *the entropy production stems solely from the diagonal part of the density operator.*

Of course, we have not proved that the van Hove limit is necessary for irreversibility. In fact, we conjecture that the partitioning of the Hamiltonian, together with some limit on λ^n, n arbitrary but finite, will do the trick. This has not as yet been proven.

6. RESULTS IN THE PRESENCE OF AN EXTERNAL FIELD: LINEAR RESPONSE FOR CLASSICAL FREQUENCIES

We continue the original goal of the paper, viz., to obtain new linear response formulas. With reference to Fig. 1, we note that the inhomogeneous master equation, path (c) of that figure, was obtained in Secs. 3 and 4. In this section and the next one we complete path (d) of that figure, i.e., we will obtain new linear response results. Thus, we will obtain the response of a quantity $B(t)$ to an external field, and determine the response function, the relaxation function, the generalized susceptibility, and the generalized conductivity. Basically, we will proceed as in Kubo's theory, though the equation for the evolution of the density operator is no longer the von Neumann equation, but the inhomogeneous master equation which resulted from the van Hove limit. Also, the $B(t)$ to be considered are the macroscopic, reduced operators $B^R(t)$. Note that $B^R = B^S = B$, but $\dot{B}^R = J_B \neq \dot{B}^S$ [see (4.28)], so the $t = 0$ operator B^R need not to be superscripted, but the operator $\dot{B}^R$ does need to be.

A. The interaction form

In the present section we restrict ourselves to linear response at classical frequencies. We can then neglect the nondiagonal contributions (except for some cases, cf. Sec. 2.A, last paragraph). Thus, we use the evolution of the diagonal density operator $\rho_d^R(t)$ in the presence of an external field perturbation, turned on at $t = 0$; the result was given by Eq. (4.33):

$$\rho_d^R(t) - \rho_{\mathrm{eq}} = \beta \int_0^t dt'\, F(t') e^{-\Lambda_d(t-t')} \rho_{\mathrm{eq}} (\dot{A}^R)_d \quad (t > 0). \quad (6.1)$$

The response of a dynamical variable $B \approx B_d$ is then

$$\langle \Delta B(t) \rangle = \mathrm{Tr}\rho_d^R(t) B_d - \mathrm{Tr}\rho_{\mathrm{eq}} B_d$$
$$= \beta \int_0^t dt'\, F(t') \mathrm{Tr}(B_d\, e^{-\Lambda_d(t-t')} \rho_{\mathrm{eq}} (\dot{A}^R)_d). \quad (6.2)$$

Formally, we write

$$\langle \Delta B(t) \rangle = \int_0^t dt'\, F(t') \phi_{BA}^{\mathrm{cl}}(t-t'), \quad (6.3)$$

where ϕ_{BA}^{cl} is the classical response function for B due to the external field Hamiltonian $-F(t)A$. Comparing Eq. (6.2) with (6.3), we see that the response function is given by

$$\phi_{BA}^{\mathrm{cl}}(t) = \beta\, \mathrm{Tr}(B_d\, e^{-\Lambda_d t} \rho_{\mathrm{eq}} (\dot{A}^R)_d). \quad (6.4)$$

We now use again Lemma 2 of Appendix C. Then Eq. (6.4) is carried over into

$$\phi_{BA}^{\mathrm{cl}}(t) = \beta\, \mathrm{Tr}(\rho_{\mathrm{eq}} (\dot{A}^R)_d e^{-\Lambda_d t} B_d). \quad (6.5)$$

Also, with

$$e^{-\Lambda_d t} B_d = B_d^R(t) \quad (6.6)$$

we have

$$\phi_{BA}^{\mathrm{cl}}(t) = \beta\, \mathrm{Tr}[\rho_{\mathrm{eq}} (\dot{A}^R)_d B_d^R(t)]. \quad (6.7)$$

We now take the Laplace transform of the convolution integral (6.3) and write

$$b(s) = \chi(s) f(s), \quad (6.8)$$

where

$$b(s) = \int_0^\infty e^{-st} \langle \Delta B(t) \rangle dt, \quad (6.9)$$

$$f(s) = \int_0^\infty e^{-st} F(t) dt, \quad (6.10)$$

$$\chi_{BA}(s) = \int_0^\infty e^{-st} \phi_{BA}^{\mathrm{cl}}(t) dt. \quad (6.11)$$

Here χ is the generalized susceptibility. In the frequency domain we have for $s = i\omega + 0$ the complex susceptibility

$$\chi_{BA}(i\omega) = \beta \int_0^\infty dt\, e^{-i\omega t} \mathrm{Tr}[\rho_{\mathrm{eq}} (\dot{A}^R)_d B_d^R(t)]. \quad (6.12)$$

While this has in appearance the same form as Kubo's result [Part I, Eq. (3.22')], the time dependence (6.6) of the reduced time dependent operators is entirely different from the time dependence of the Heisenberg operators occurring in the Kubo formulas; Eq. (6.12) converges and can, in fact, be expressed in the resolvent of the master operator Λ_d:

$$\chi_{BA}(i\omega) = \beta \,\mathrm{Tr}\{\rho_{\mathrm{eq}}(\dot{A}^R)_d [i\omega + \Lambda_d]^{-1} B_d\}. \quad (6.13)$$

As in Kubo's linear response theory, we can also seek the response of a quantity $\dot{B}$. Analogous to Eq. (6.2) we find

$$\langle \Delta \dot{B}(t)\rangle = \beta \int_0^t dt' \, F(t') \mathrm{Tr}[(\dot{B}^R)_d e^{-\Lambda_d (t-t')} \rho_{\mathrm{eq}} (\dot{A}^R)_d]$$
$$= \beta \int_0^t dt' \, F(t') \mathrm{Tr}[\rho_{\mathrm{eq}} (\dot{A}^R)_d e^{-\Lambda_d (t-t')} (\dot{B}^R)_d]. \quad (6.14)$$

We now introduce the relaxation function. It will be defined by

$$\Psi^{\mathrm{cl}}_{BA}(t) = \phi^{\mathrm{cl}}_{BA}(t). \quad (6.15)$$

{To prove that this function gives the relaxation after a force is turned off, we must establish that $\Psi^{\mathrm{cl}}_{BA}(t)$ is the time integral of $\phi^{\mathrm{cl}}_{BA}(t)$ [see Part I, Eqs. (2.17) and (2.18)].} With Eq. (6.15), the expression of Eq. (6.14) can be written as

$$\langle \Delta \dot{B}(t)\rangle = \int_0^t dt' \, F(t') \Psi^{\mathrm{cl}}_{B\dot{A}}(t - t'), \quad (6.16)$$

where

$$\Psi^{\mathrm{cl}}_{B\dot{A}}(t) = \beta \,\mathrm{Tr}[\rho_{\mathrm{eq}} (\dot{A}^R)_d e^{-\Lambda_d t} (\dot{B}^R)_d]. \quad (6.17)$$

We take again the Laplace transform of Eq. (6.14), which leads to

$$\dot{b}(s) = L(s) f(s), \quad (6.18)$$

where $\dot{b}(s)$ is the Laplace transform of $\langle \Delta \dot{B}(t)\rangle$, and where

$$L_{BA}(s) = \int_0^\infty e^{-st} \Psi^{\mathrm{cl}}_{B\dot{A}}(t) dt \quad (6.19)$$

is the generalized conductivity. In the frequency domain we have for $s = i\omega + 0$ the complex conductivity

$$L_{BA}(i\omega) = \beta \int_0^\infty dt \, e^{-i\omega t} \mathrm{Tr}[\rho_{\mathrm{eq}} (\dot{A}^R)_d (\dot{B}^R)_d(t)]. \quad (6.20)$$

This can also be expressed in the resolvent

$$L_{BA}(i\omega) = \beta \,\mathrm{Tr}\{\rho_{\mathrm{eq}} (\dot{A}^R)_d [i\omega + \Lambda_d]^{-1} (\dot{B}^R)_d\}. \quad (6.21)$$

Equations (6.13) and (6.21) are the main new results.

We make some additional remarks. First, it is noted that the equilibrium expectation value of a flux $\dot{A}^R(t)$ is zero, which is the same feature as in the Heisenberg description [Part I, Eq. (3.24)]. The proof for the present case is as follows: we have, see Eq. (4.28)

$$\langle \dot{A}^R(t)\rangle_{\mathrm{eq}} = -\mathrm{Tr}(\rho_{\mathrm{eq}} \Lambda_d e^{-\Lambda_d t} A_d)$$
$$+ (1/z_0) \mathrm{Tr}[e^{-\beta H^0} e^{-\Lambda_d t} (\dot{A})_d]$$
$$= -\mathrm{Tr}[(e^{-\Lambda_d t} A_d) \Lambda_d \rho_{\mathrm{eq}}]$$
$$+ (1/z_0 h i) \mathscr{P} \,\mathrm{Tr}\{A e^{-\Lambda_d t} [H^0, e^{-\beta H''}]\} = 0, \quad (6.22)$$

where we used Lemmas 1 and 2 of Appendix C, and the result

$$\Lambda_d \rho_{\mathrm{eq}} = \frac{2\pi\lambda^2}{\hbar} \sum_{\gamma\gamma''} |\gamma\rangle \langle\gamma| \langle\gamma| V |\gamma''\rangle|^2$$
$$\times \delta(\epsilon_\gamma - \epsilon_{\gamma''})(e^{-\beta\epsilon_{\gamma''}} - e^{-\beta\epsilon_\gamma})/z^0 = 0. \quad (6.23)$$

A fortiori,

$$\langle \dot{A}^R\rangle_{\mathrm{eq}} = \lim_{t\to\infty} \langle \dot{A}^R(t)\rangle_{\mathrm{eq}} = 0. \quad (6.22')$$

Defining now $\Delta\dot{A}^R(t) = \dot{A}^R(t) - I\langle \dot{A}^R\rangle_{\mathrm{eq}}$, we see that at all times $\Delta\dot{A}^R(t) = \dot{A}^R(t)$.

Finally, we will apply Theorem 3 [Eq. (2.12)] to the reduced time dependent operators. This gives

$$B^R_d(t) = \sum_\gamma |\gamma\rangle \langle\gamma| e^{-M_\gamma t} \langle\gamma| B_d |\gamma\rangle, \quad (6.24)$$

where M is the master operator in function space given in Eq. (2.13). Equations (6.13) and (6.21) are now carried over in

$$\chi_{BA}(i\omega) = \beta \sum_\gamma [p_{\mathrm{eq}}(\gamma) \langle\gamma| J_A |\gamma\rangle (i\omega + M_\gamma)^{-1} \langle\gamma| B |\gamma\rangle], \quad (6.25)$$

$$L_{BA}(i\omega)$$
$$= \beta \sum_\gamma [p_{\mathrm{eq}}(\gamma) \langle\gamma| J_A |\gamma\rangle (i\omega + M_\gamma)^{-1} \langle\gamma| J_B |\gamma\rangle]. \quad (6.26)$$

This gives χ and L in terms of the resolvent of the functional master operator M.

B. The Schrödinger form

We will obtain another result in which the time dependence is vested in $\rho^R(t)$ rather than in $B^R(t)$ or $\dot{B}^R(t)$. To this purpose we solve directly Eq. (4.31). We define the Green's function $g(\gamma,t;\gamma',t')$ of the master operator $(\partial/\partial t) + M$ in function space:

$$\frac{\partial}{\partial t} g(\gamma,t;\gamma',t') + M g(\gamma,t;\gamma',t') = \delta(\gamma - \gamma')\delta(t - t'). \quad (6.27)$$

The solution of Eq. (4.31) is then in terms of this Green's function:

$$p^R(\gamma,t) = p_{\mathrm{eq}}(\gamma) + \beta \int_0^{t+0} dt'$$
$$\times \int \Delta\gamma' F(t') p_{\mathrm{eq}}(\gamma') g(\gamma,t;\gamma',t') \langle\gamma'| \dot{A}^R |\gamma'\rangle. \quad (6.28)$$

As in Part I, $\Delta\gamma$ denotes $Z(\gamma)d\gamma$, where $Z(\gamma)$ is the density of states for the quasicontinuous quantum variables γ. We still write $\langle\gamma'|\dot{A}^R|\gamma'\rangle = \dot{A}^R_{\gamma'}$. For the response of a variable B we obtain, writing also $\langle\gamma|B|\gamma\rangle = B_\gamma$,

$$\langle \Delta B(t)\rangle = \sum_\gamma [p^R(\gamma,t) B_\gamma - p_{\mathrm{eq}}(\gamma) B_\gamma]$$
$$= \int \Delta\gamma [p^R(\gamma,t) - p_{\mathrm{eq}}(\gamma)] B_\gamma$$
$$= \beta \int_0^{t+0} dt' \int\int \Delta\gamma \Delta\gamma' F(t') p_{\mathrm{eq}}(\gamma')$$
$$\times g(\gamma,t;\gamma',t') \dot{A}^R_{\gamma'} B_\gamma. \quad (6.29)$$

We note that $g(\gamma,t;\gamma',t') = g(\gamma,t-t';\gamma',0)$ [cf. Eq. (6.27)]. Thus, Eq. (6.29) is again a convolution integral. The response function is therefore

$$\phi^{\mathrm{cl}}_{BA}(t) = \beta \int\int \Delta\gamma \Delta\gamma' p_{\mathrm{eq}}(\gamma') g(\gamma,t;\gamma',0) \dot{A}^R_{\gamma'} B_\gamma. \quad (6.30)$$

The complex susceptibility is the one-sided Fourier transform of ϕ, i.e.,

$$\chi_{BA}(i\omega) = \beta \int_0^\infty dt\, e^{-i\omega t} \int\int \Delta\gamma \Delta\gamma' p_{\rm eq}(\gamma') g(\gamma,t;\gamma',0) \times \dot{A}^R_\gamma B_\gamma. \tag{6.31}$$

Let $G(\gamma,s;\gamma')$ be the Laplace transform of $g(\gamma,t;\gamma',0)$. Then, Eq. (6.31) results in

$$\chi_{BA}(i\omega) = \beta \int\int \Delta\gamma \Delta\gamma' p_{\rm eq}(\gamma') G(\gamma,i\omega;\gamma') \dot{A}^R_\gamma B_\gamma. \tag{6.32}$$

In a similar way we could have obtained the response of a quantity $\dot{B}$, which leads to

$$L_{BA}(i\omega) = \beta \int\int \Delta\gamma \Delta\gamma' p_{\rm eq}(\gamma') G(\gamma,i\omega;\gamma') \dot{A}^R_\gamma \dot{B}^R_\gamma. \tag{6.33}$$

The results (6.32) and (6.33) could also have been obtained from the previous results (6.25) and (6.26), for the resolvent $(s+M)^{-1}$ can be shown to be the integral operator associated with the Green's function $G(\gamma,s;\gamma')$:

$$(s+M_\gamma)^{-1} f(\gamma) = \int \Delta\gamma' G(\gamma,s;\gamma') f(\gamma'). \tag{6.34}$$

Finally, we put the expressions for χ and L in a form which shows the stochastic nature of our present results. We recall that the Green's function g is also equal to the conditional probability of the homogeneous master equation [see Part I, Eq. (7.12)]:

$$g(\gamma,t;\gamma',0) = P_{\rm sp}(\gamma,t\,|\gamma',0). \tag{6.35}$$

We now introduce the two-point probability

$$W_{2,\rm sp}(\gamma,t;\gamma',0) = P_{\rm sp}(\gamma,t\,|\gamma',0) p_{\rm eq}(\gamma'). \tag{6.36}$$

Then, Eqs. (6.32) and (6.33) read also

$$\chi_{BA}(i\omega) = \beta \int_0^\infty dt\, e^{-i\omega t} \int\int \Delta\gamma \Delta\gamma' W_{2,\rm sp}(\gamma,t;\gamma',0) \dot{A}^R_\gamma B_\gamma \tag{6.37}$$

and

$$L_{BA}(i\omega) = \beta \int_0^\infty dt\, e^{-i\omega t} \int\int \Delta\gamma \Delta\gamma' W_{2,\rm sp}(\gamma,t;\gamma',0) \dot{A}^R_\gamma \dot{B}^R_\gamma, \tag{6.38}$$

respectively. The $W_{2,\rm sp}$ introduced here is the standard two-point probability which measures the deviations from the equilibrium state, as we show explicitly in the next subsection.

C. The correlation function

The correlation function for fluctuations from the equilibrium state of two variables $\mathscr{C}$ and $\mathscr{D}$ is defined by

$$\Phi_{CD}(t) \equiv \langle \Delta\mathscr{C}(t) \Delta\mathscr{D}(0) \rangle_{\rm eq} = \tfrac{1}{2} \langle [\Delta C^R(t), \Delta D]_+ \rangle_{\rm eq} = \tfrac{1}{2} \{ {\rm Tr}[\rho_{\rm eq} \Delta C^R(t) \Delta D] + {\rm Tr}[\rho_{\rm eq} \Delta D \Delta C^R(t)] \}. \tag{6.39}$$

In the classical frequency range C and D can be replaced by C_d and D_d, respectively, and the anticommutator is of no effect; thus,

$$\Phi^{\rm cl}_{CD}(t) = {\rm Tr}[\rho_{\rm eq} (e^{-M_d t} \Delta C_d) \Delta D_d]. \tag{6.40}$$

We now use Theorem 3 [Eq. (2.12)]; this yields, if we evaluate the trace in the representation $\{|\gamma\rangle\}$,

$$\Phi^{\rm cl}_{CD}(t) = {\rm Tr}\Big[\rho_{\rm eq} \sum_\gamma |\gamma'\rangle \langle\gamma'| (e^{-M_\gamma t} \langle\gamma'|\Delta C_d|\gamma'\rangle) \Delta D_d\Big] = \sum_\gamma p_{\rm eq}(\gamma) (e^{-M_\gamma t} \langle\gamma|\Delta C_d|\gamma\rangle) \langle\gamma|\Delta D_d|\gamma\rangle. \tag{6.41}$$

It is readily shown that $\exp(-M_\gamma t)$ is a Green's integral operator

$$e^{-M_\gamma t} f(\gamma) = \int \Delta\gamma' g(\gamma,t;\gamma',0) f(\gamma') \tag{6.42}$$

[note that Eq. (6.42) is the Laplace inverse of Eq. (6.34)]. Thus, Eq. (6.41) yields,

$$\Phi^{\rm cl}_{CD}(t) = \int\int \Delta\gamma \Delta\gamma' p_{\rm eq}(\gamma) g(\gamma,t;\gamma',0) \Delta C_\gamma \Delta D_\gamma = \int\int \Delta\gamma \Delta\gamma' W_{2,\rm sp}(\gamma',t;\gamma,0) \Delta C_\gamma \Delta D_\gamma \tag{6.43}$$

[where we still used the symmetry property $g(\gamma,t;\gamma',0) = g(\gamma',t;\gamma,0)$]. Equation (6.43) is the standard definition of a two-point correlation function in stochastic theory. We thus showed full equivalence of the van Hove limit form or reduced Heisenberg form (6.40) with the stochastic form or Schrödinger form (6.43). This development parallels the procedure of Part I, last paragraph of Sec. 9.2. However, in the present discussion we stayed within the subdynamics of H^0.

7. THE LINEAR RESPONSE RESULT FOR QUANTUM FREQUENCIES

We give some brief results for the linear response coefficients L and χ for the case that the nondiagonal correlations are included. The solution of the kinetic equation for the full ρ, [Eq. (4.41)], reads

$$\rho^R(t) = \rho_{\rm eq} + u(t) \int_0^t dt'\, F(t') \times e^{-(\Lambda_d + i\mathscr{L}^0)(t-t')} \int_0^\beta d\beta'\, \rho_{\rm eq} e^{\hbar\beta' \mathscr{L}^0} \dot{A}^R. \tag{7.1}$$

Again we look for the response of an operator B:

$$\langle \Delta B(t) \rangle = \int_0^t dt'\, F(t') {\rm Tr}\Big(B\, e^{-(\Lambda_d + i\mathscr{L}^0)(t-t')} \times \int_0^\beta d\beta'\, \rho_{\rm eq} e^{\hbar\beta' \mathscr{L}^0} \dot{A}^R \Big); \tag{7.2}$$

this leads to the response function

$$\phi_{BA}(t) = \int_0^\beta d\beta'\, {\rm Tr}\big(B\, e^{-(\Lambda_d + i\mathscr{L}^0)t} \rho_{\rm eq} e^{\hbar\beta' \mathscr{L}^0} \dot{A}^R\big). \tag{7.3}$$

Next we use Lemma 5 of Appendix C. We then obtain the form

$$\phi_{BA}(t) = \int_0^\beta d\beta'\, {\rm Tr}[\rho_{\rm eq} (e^{\hbar\beta' \mathscr{L}^0} \dot{A}^R) e^{-(\Lambda_d - i\mathscr{L}^0)t} B]. \tag{7.4}$$

This is the interaction form. We can also split it into diagonal and nondiagonal contributions. Using Eq. (5.7), we have

$$B^R(t) = e^{-(\Lambda_d - i\mathscr{L}^0)t}B = e^{-\Lambda_d t}B_d + e^{i\mathscr{L}^0 t}B_{nd} = B^R_d(t) + B^R_{nd}(t), \tag{7.5}$$

where

$$B^R_{nd}(t) \equiv B^I_{nd}(t), \quad \text{while also} \quad (\dot{B}^R)_{nd} \equiv (\dot{B}^I)_{nd}. \tag{7.6}$$

Further,

$$e^{\hbar\beta'\mathscr{L}^0}\dot{A}^R = (\dot{A}^R)_d + (\dot{A}^R(-i\hbar\beta'))_{nd}. \tag{7.7}$$

Hence, as an alternate form for ϕ we obtain

$$\phi_{BA}(t) = \beta \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_d B^R_d(t)] + \int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta')B^R_{nd}(t)]. \tag{7.8}$$

Thus, *ϕ is the sum of the diagonal part* (6.7) *plus a Kubo expression for the nondiagonal part.* Likewise for the relaxation function,

$$\Psi_{BA}(t) = \beta \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_d(\dot{B}^R)_d(t)] + \int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta')(\dot{B}^R)_{nd}(t)]. \tag{7.9}$$

For χ and L we have the following general formulas, based on the compact form (7.4):

$$\chi_{BA}(i\omega) = \int_0^\infty dt\, e^{-i\omega t}\int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(e^{\hbar\beta'\mathscr{L}^0}\dot{A}^R) \times e^{-(\Lambda_d - i\mathscr{L}^0)t}B] = \int_0^\beta d\beta' \operatorname{Tr}\left[\rho_{eq}(e^{\hbar\beta'\mathscr{L}^0}\dot{A}^R) \times \frac{1}{\Lambda_d - i\mathscr{L}^0 + i\omega}B\right], \tag{7.10}$$

$$L_{BA}(i\omega) = \int_0^\infty dt\, e^{-i\omega t}\int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(e^{\hbar\beta'\mathscr{L}^0}\dot{A}^R) \times e^{-(\Lambda_d - i\mathscr{L}^0)t}\dot{B}^R] = \int_0^\beta d\beta' \operatorname{Tr}\left[\rho_{eq}(e^{\hbar\beta'\mathscr{L}^0}\dot{A}^R) \times \frac{1}{\Lambda_d - i\mathscr{L}^0 + i\omega}\dot{B}^R\right]. \tag{7.11}$$

We also have the more explicit form based on Eqs. (7.8) and (7.9):

$$\chi_{BA}(i\omega) = \int_0^\infty dt\, e^{-i\omega t}\left\{\beta \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_d B^R_d(t)] + \int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta')B^R_{nd}(t)]\right\}$$
$$= \beta \operatorname{Tr}\left[\rho_{eq}(\dot{A}^R)_d \frac{1}{\Lambda_d + i\omega}B_d\right] + \int_0^\beta d\beta' \operatorname{Tr}\left[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta') \times \frac{1}{-i\mathscr{L}^0 + i\omega}B_{nd}\right] + \pi\int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta') \times \delta(\omega - \mathscr{L}^0)B_{nd}], \tag{7.12}$$

$$L_{BA}(i\omega) = \int_0^\infty dt\, e^{-i\omega t}\left\{\beta \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_d(\dot{B}^R)_d(t)] + \int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta')(\dot{B}^R)_{nd}(t)]\right\}$$
$$= \beta \operatorname{Tr}\left[\rho_{eq}(\dot{A}^R)_d \frac{1}{\Lambda_d + i\omega}(\dot{B}^R)_d\right] + \int_0^\beta d\beta' \operatorname{Tr}\left[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta') \times \frac{1}{-i\mathscr{L}^0 + i\omega}(\dot{B}^R)_{nd}\right] + \pi\int_0^\beta d\beta' \operatorname{Tr}[\rho_{eq}(\dot{A}^R)_{nd}(-i\hbar\beta') \times \delta(\omega - \mathscr{L}^0)(\dot{B}^R)_{nd}]. \tag{7.13}$$

The Schrödinger forms are obtained as in the previous section. The following result is then found:

$$\chi_{BA} = \beta\int\int \Delta\gamma\Delta\gamma'\rho_{eq}(\gamma)G(\gamma',i\omega;\gamma)\langle\gamma|\dot{A}^R|\gamma\rangle\langle\gamma'|B|\gamma'\rangle + \hbar\pi\int\int\Delta\gamma\Delta\gamma'\frac{p_{eq}(\gamma) - p_{eq}(\gamma')}{\epsilon_\gamma - \epsilon_\gamma}\delta(\hbar\omega + \epsilon_\gamma - \epsilon_{\gamma'}) \times \langle\gamma|\dot{A}^R|\gamma'\rangle\langle\gamma'|B|\gamma\rangle + \hbar i\int\int\Delta\gamma\Delta\gamma'\frac{p_{eq}(\gamma) - p_{eq}(\gamma')}{\epsilon_\gamma - \epsilon_\gamma} \times \frac{1}{-\hbar\omega + (\epsilon_{\gamma'} - \epsilon_\gamma)}\langle\gamma|\dot{A}^R|\gamma'\rangle\langle\gamma'|B|\gamma\rangle, \tag{7.14}$$

where integrals involve the Cauchy principal value. A similar result holds for L_{BA} with $\dot{B}^R$ replacing B.

Finally, we make some remarks on the fluctuation–dissipation theorem. The correlation function (6.39) splits into two parts

$$\Phi_{CD}(t) = \operatorname{Tr}[\rho_{eq}\Delta C^R_d(t)\Delta D_d] + \tfrac{1}{2}\operatorname{Tr}\{\rho_{eq}[\Delta C^R_{nd}(t),\Delta D_{nd}]_+\}, \tag{7.14}$$

where $[\,]_+$ is the anticommutator. These functions can now be coupled with the relaxation function or with the response function. Using the same manipulations as in Part I, one obtains the following four fluctuation–dissipation theorems:

$$S^d_{BA}(\omega) = -4kT\frac{1}{\omega}\{[\chi^{\cdot d}_{BA}(\omega)]^s - i[\chi^{\cdot d}_{BA}(\omega)]^a\}, \tag{7.15}$$

$$S^{nd}_{BA}(\omega) = -4\mathscr{E}(\omega,T)\frac{1}{\omega}\{[\chi^{\cdot nd}_{BA}(\omega)]^s - i[\chi^{\cdot nd}_{BA}(\omega)]^a\}, \tag{7.16}$$

$$S^d_{BA}(\omega) = 4kT\{[L^{\cdot d}_{BA}(\omega)]^s + i[L^{\cdot d}_{BA}(\omega)]^a\}, \tag{7.17}$$

$$S^{nd}_{BA}(\omega) = 4\mathscr{E}(\omega,T)\{[L^{\cdot nd}_{BA}(\omega)]^s + i[L^{\cdot nd}_{BA}(\omega)]^a\}. \tag{7.18}$$

It is evident that the quantum correction factor occurs only in the nondiagonal contributions. Thus, the total spectra do not satisfy the exact fluctuation–dissipation theorem. This can be seen as a flaw of the van Hove limit results. Two notes are in order, however, to understand the nature of this flaw. First, the van Hove limit required that the times were asymptotically large, or the frequencies small. Therefore, the results must be stretched in order to be valid at quantum frequencies. We also note the approximation made in Eq. (4.20). Secondly, we observe that the diagonal spectra go as $\|(\Lambda_d \pm i\omega)^{-1}\|$, which goes to zero for very high frequencies. Thus, one can expect that at quantum frequencies near-

ly all the noise stems form the nondiagonal parts. In this sense, then, the usual fluctuation–dissipation theorem with inclusion of the quantum correction factor is approximately confirmed.

8. DERIVATION OF A QUANTUM BOLTZMANN EQUATION

The Boltzmann equation, dealing with the kinetic development of the one-particle distribution function, can be obtained from the N-particle Liouville equation, by integration over the coordinates and momenta of $N-1$ particles. The classical derivation of H. Grad is well known.[16] For Boltzmann's own derivation and comments on the validity of it we refer to his book[17]. Since the Boltzmann equation is irreversible, it is clear that somewhere along the way a randomness assumption must be introduced. Our point of view is that such a randomness assumption is best made at the many-body level, such as we did in this paper. The many-body equation which then is the point of departure is the master equation. The first moment equation of the master equation is the Boltzmann equation. This was shown in principle by van Hove[5], but because of his restricted master equation, the streaming terms were absent. Using our general inhomogeneous master equation, we will show that a full Boltzmann equation is easily obtained, without severe additional assumptions; or, for that matter, we could derive a full hierarchy of moment equations, analogous to the BBGKY hierarchy, but we hope to do that elsewhere. The only assumption is a truncation rule. We assume that, in the kinetic equation for a K-particle function, a $K+1$ particle function can be factored. In particular for the present we assume that, in the equation for $\langle n_i \rangle$,

$$\langle n_i n_j \rangle = \langle n_i \rangle \langle n_j \rangle + \langle \Delta n_i \Delta n_j \rangle \approx \langle n_i \rangle \langle n_j \rangle \quad (i \neq j). \tag{8.1}$$

This amounts to the neglect of cross correlation, which is correct in a grand canonical ensemble but not exact in a canonical ensemble, though the correction is of order $\Sigma(n_i)^{-1}$, i.e., negligible for large systems.

Since our master equation is in quantum mechanical form, the Boltzmann equation to be derived pertains to a distribution over one particle *quantum states*. As we do not in this article wish to enter into quantum phase space analogs (Wigner distribution), our quantum Boltzmann equation is only valid for homogeneous systems; thus, the spatial gradient term does not occur. On the other hand, the result to be derived is not tied to a momentum state or semiclassical $\mathbf{k}$-space description. The force streaming terms which we obtain are fully quantum mechanical. The equation can thus be applied to situations which can not be described by the standard Boltzmann equation, such as transverse magnetic resistance (Landau levels)[18], or conduction in a MOS field-effect transistor, which shows discrete transverse energy quantization under strong inversion[19]. A study of these applications is being undertaken.

In the derivation we must specify the interactions λV. Two cases will be considered. Two-body interactions of particles with mass (fermion–boson or boson–boson) is dealt with in Sec. 8.A. For electrons in a solid, electron–phonon interaction is taken up in Sec. 8.B. The derivations go most smoothly by using the formalism of second quantization.

A. Binary interactions of particles with mass

Since we deal with two kinds of particles which only weakly interact (so we assume) with partners of their own kind, the Hamiltonian H^0 commutes with the occupancy operators of the quantum states of each kind. We will first consider a system composed of fermions and bosons, with occupancy operators $\mathbf{n}_\xi$ and $\mathbf{N}_\eta$, respectively; here $|\xi)$ refers to an ordered set of the one-particle fermion states and $|\eta)$ denotes a one-particle boson state. The many-body states of H^0 are taken as

$$|\gamma\rangle = |\{n_\xi\},\{N_\eta\}\rangle, \tag{8.2}$$

where n_ξ and N_η are occupancy numbers, while H^0 is written as

$$\begin{aligned} H^0 &= \sum_\xi \mathbf{n}_\xi e_\xi + \sum_\eta \mathbf{N}_\eta E_\eta \\ &= \sum_\xi c_\xi^\dagger c_\xi e_\xi + \sum_\eta a_\eta^\dagger a_\eta E_\eta; \end{aligned} \tag{8.3}$$

here e_ξ and E_η are one-particle energy eigenvalues, and the c's and a's are raising and lowering operators. The total Hamiltonian is again $H = H^0 + \lambda V$, with the binary interaction

$$H^1 = \lambda V = \sum_{\xi''\xi'\eta''\eta'} (\xi''\eta''|\lambda v|\xi'\eta') c_{\xi''}^\dagger a_{\eta''}^\dagger a_{\eta'} c_{\xi'}, \tag{8.4}$$

where $v = v(\mathbf{r} - \mathbf{R})$ is the two-body interaction potential. (Note the we use round bracket kets $|)$ for states pertaining to one particle of each or both kinds, while we use angular kets $|\rangle$ for the many-body states.)

As an example, consider the case of free particles treated by van Hove[5]. Then $|\xi) \equiv |\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{r}}/\Omega^{1/2}$ and $|\eta) \equiv |\mathbf{K}) = e^{i\mathbf{K}\cdot\mathbf{R}}/\Omega^{1/2}$, where Ω is the volume of the system. We have

$$\begin{aligned} &(\mathbf{k}''\mathbf{K}''|\lambda v|\mathbf{k}'\mathbf{K}') \\ &= \Omega^{-2} \int\int e^{-i\mathbf{k}''\cdot\mathbf{r}} e^{-i\mathbf{K}''\cdot\mathbf{R}} v(\mathbf{r}-\mathbf{R}) e^{i\mathbf{k}'\cdot\mathbf{r}} e^{i\mathbf{K}'\cdot\mathbf{R}} d^3r\, d^3R. \end{aligned} \tag{8.5}$$

Let

$$v(\mathbf{r}-\mathbf{R}) = \Omega^{-1} \sum_{\mathbf{q}} e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{R})} v_{\mathbf{q}}; \tag{8.6}$$

then

$$(\mathbf{k}''\mathbf{K}''|\lambda v|\mathbf{k}'\mathbf{K}') = \Omega^{-1} \sum_{\mathbf{q}} \delta_{\mathbf{k}'',\mathbf{k}'+\mathbf{q}} \delta_{\mathbf{K}'',\mathbf{K}'-\mathbf{q}} v_{\mathbf{q}} \tag{8.7}$$

so that

$$\begin{aligned} V &= \sum_{\mathbf{k}''\mathbf{k}'\mathbf{K}''\mathbf{K}'\mathbf{q}} \Omega^{-1} \delta_{\mathbf{k}'',\mathbf{k}'+\mathbf{q}} \delta_{\mathbf{K}'',\mathbf{K}'-\mathbf{q}} v_{\mathbf{q}} c_{\mathbf{k}''}^\dagger c_{\mathbf{k}'} a_{\mathbf{K}''}^\dagger a_{\mathbf{K}'} \\ &= \Omega^{-1} \sum_{\mathbf{k}'\mathbf{K}'\mathbf{q}} v_{\mathbf{q}} c_{\mathbf{k}'+\mathbf{q}}^\dagger c_{\mathbf{k}'} a_{\mathbf{K}'-\mathbf{q}}^\dagger a_{\mathbf{K}'}, \end{aligned} \tag{8.8}$$

in accord with van Hove.

We continue with the general case given by Eq. (8.4). We must compute the matrix elements $\langle\bar{\gamma}|\lambda V|\gamma\rangle$ for the transition probabilities of the master equation. To that end we need the raising and lowering rules in the following form:

$$a^\dagger_{\eta''}|\{\ \},\{N_\eta\}\rangle = \sqrt{N_{\eta''}+1}\,|\{\ \},...,N_{\eta''}+1,...\rangle, \tag{8.9}$$

$$a_{\eta'}|\{\ \},\{N_\eta\}\rangle = \sqrt{N_{\eta'}}\,|\{\ \},...,N_{\eta'}-1,...\rangle, \tag{8.10}$$

$$c^\dagger_{\xi''}|\{n_\xi\},\{\ \}\rangle = (-1)^{\Sigma(1,\xi''-1)}\sqrt{1-n_{\xi''}} \times |...,1-n_{\xi''},...,\{\ \}\rangle, \tag{8.11}$$

$$c_{\xi'}|\{n_\xi\},\{\ \}\rangle = (-1)^{\Sigma(1,\xi'-1)}\sqrt{n_{\xi'}} \times |...,1-n_{\xi'},...,\{\ \}\rangle. \tag{8.12}$$

Here $\Sigma(1,\xi)$ is the sum of all occupancies of states 1,2,...,ξ. There is a slight problem with the fermion operators, for which the square roots are often omitted, since for the occupancy operators $\mathbf{n}_\xi = \mathbf{n}_\xi^2$. For the present derivation it is desirable to maintain the square roots. This is briefly discussed in Appendix D. The occupancy numbers are $n_\xi = 0$ or 1; thus, Eq. (8.11) yields only nonzero for $n_{\xi''} = 0$, so that $1-n_{\xi''} = 1$ denotes a raising, while Eq. (8.12) yields only nonzero for $n_{\xi'} = 1$, so that $1-n_{\xi'} = 0$ denotes a lowering. From the above rules we find that the matrix element $\langle\bar\gamma|\lambda V|\gamma\rangle = \langle\{\bar n_\xi\},\{\bar N_\eta\}|\lambda V|\{n_\xi\},\{N_\eta\}\rangle$ is nonzero only if for given ξ'',ξ',η'', and η', and for given $|\gamma\rangle$, we take $|\bar\gamma\rangle$ such that

$$\left.\begin{aligned}&\bar n_\xi = n_\xi, \quad \text{for } \xi\neq\xi' \text{ or } \xi'',\\ &\bar n_{\xi''} = 1-n_{\xi''} \quad \text{and} \quad \bar n_{\xi'} = 1-n_{\xi'}\end{aligned}\right\} \tag{8.13}$$

and

$$\left.\begin{aligned}&\bar N_\eta = N_\eta, \quad \text{for } \eta\neq\eta' \text{ or } \eta'',\\ &\bar N_{\eta''} = N_{\eta''}+1 \quad \text{and} \quad \bar N_{\eta'} = N_{\eta'}-1\end{aligned}\right\}; \tag{8.14}$$

the latter statement can also be summarized by

$$\bar N_\eta + \delta_{\eta\eta'} = N_\eta + \delta_{\eta\eta''}. \tag{8.14a}$$

(We still note that always $\xi''\neq\xi'$ and $\eta''\neq\eta'$, since λV is nondiagonal.) Thus, with each choice of ξ'',ξ',η'',η' corresponds only one state $|\bar\gamma\rangle = |\bar\gamma_{\xi''\xi'\eta''\eta'}\rangle$, fixed by Eqs. (8.13) and (8.14). The value of the matrix element for the connected states $|\bar\gamma_{\xi''\xi'\eta''\eta'}\rangle$ and $|\gamma\rangle$ is

$$\begin{aligned}&\langle\bar\gamma_{\xi''\xi'\eta''\eta'}|\lambda V|\gamma\rangle\\ &= (-1)^{\Sigma(1,\xi''-1)}(-1)^{\Sigma(1,\xi'-1)}\\ &\quad\times(\xi''\eta''|\lambda v|\xi'\eta')[(1-n_{\xi''})n_{\xi'}(1+N_{\eta''})N_{\eta'}]^{1/2},\end{aligned} \tag{8.15}$$

since all other terms of the series (8.4) yield zero. The energy difference involved in this transition is

$$\delta(\epsilon_{\bar\gamma_{\xi''\xi'\eta''\eta'}} - \epsilon_\gamma) = \delta(e_{\xi''} - e_{\xi'} + E_{\eta''} - E_{\eta'}). \tag{8.16}$$

For the transition probability we obtain, from Eqs. (1.3), (8.15), and (8.16),

$$\begin{aligned}W_{\gamma,\bar\gamma_{\xi''\xi'\eta''\eta'}} = \frac{2\pi\lambda^2}{\hbar}\,&|(\xi''\eta''|v|\xi'\eta')|^2(1-n_{\xi''})n_{\xi'}\\ &\times(1+N_{\eta''})N_{\eta'}\delta(e_{\xi''} - e_{\xi'} + E_{\eta''} - E_{\eta'}).\end{aligned} \tag{8.17}$$

Let now

$$\begin{aligned}Q(\xi''\eta'';\xi'\eta') = \frac{2\pi\lambda^2}{\hbar}\,&|(\xi''\eta''|v|\xi'\eta')|^2\\ &\times\delta(e_{\xi''} - e_{\xi'} + E_{\eta''} - E_{\eta'})\end{aligned} \tag{8.18}$$

be the two-particle scattering probability per unit time. Then

$$W_{\gamma,\bar\gamma_{\xi''\xi'\eta''\eta'}} = Q(\xi''\eta'';\xi'\eta')(1-n_{\xi''})n_{\xi'}(1+N_{\eta''})N_{\eta'}. \tag{8.19}$$

The master operator expression $Mp(\gamma,t)$ of Eq. (4.32) now becomes, with $W_{\gamma\bar\gamma} = W_{\bar\gamma\gamma}$,

$$\begin{aligned}Mp(\gamma,t) &= \sum_{\bar\gamma} W_{\gamma\bar\gamma}[p(\gamma,t) - p(\bar\gamma,t)] = \sum_{\xi''\xi'\eta''\eta'} W_{\gamma,\bar\gamma_{\xi''\xi'\eta''\eta'}}[p(\{n_\xi\},\{N_\eta\},t) - p(\{\bar n_\xi\},\{\bar N_\eta\},t)]\\ &= \sum_{\xi''\xi'\eta''\eta'} Q(\xi''\eta'';\xi'\eta')(1-n_{\xi''})n_{\xi'}(1+N_{\eta''})N_{\eta'}\,[p(\{n_\xi\},\{N_\eta\},t) - p(...,1-n_{\xi'},...,1-n_{\xi''},\\ &\quad ...,...,N_{\eta'}-1,...,N_{\eta''}+1,...,t)],\end{aligned} \tag{8.20}$$

where we substituted for $\{\bar n_\xi\}$ and $\{\bar N_\eta\}$ from Eqs. (8.13) and (8.14). We now make the adiabatic assumption that the bosons are close to thermal equilibrium at all times since we are intersted in the distribution of the fermions alone. Thus let,

$$\not{p}(\{n_\xi\},t) = \sum_{\{N_\eta\}} p(\{n_\xi\},\{N_\eta\},t), \tag{8.21}$$

while also due to the adiabatic assumption

$$p(\{n_\xi\},\{N_\eta\},t) = P(\{N_\eta\},t\,|\{n_\xi\},t)\not{p}(\{n_\xi\},t) \simeq P_{\rm eq}(\{N_\eta\})\not{p}(\{n_\xi\},t), \tag{8.22}$$

so that

$$\sum_{\{N_\eta\}} (1+N_{\eta''})N_{\eta'}p(\{n_\xi\},\{N_\eta\},t) = \langle(1+N_{\eta''})N_{\eta'}\rangle_{\rm eq}\not{p}(\{n_\xi\},t) \tag{8.23}$$

and

$$\sum_{\{N_\eta\}} (1+N_{\eta''})N_{\eta'}p(\{\bar n_\xi\},...,N_{\eta'}-1,...,N_{\eta''}+1,...,t) = \langle N_{\eta''}(1+N_{\eta'})\rangle_{\rm eq}\not{p}(\{\bar n_\xi\},t). \tag{8.24}$$

We then find

$$\begin{aligned}\sum_{\{N_\eta\}} Mp(\gamma,t) = \sum_{\xi''\xi'\eta''\eta'} &Q(\xi''\eta'';\xi'\eta')(1-\eta_{\xi''})n_{\xi'}\,[\langle(1+N_{\eta''})N_{\eta'}\rangle_{\rm eq}\not{p}(\{n_\xi\},t) - \langle N_{\eta''}(1+N_{\eta'})\rangle_{\rm eq}\\ &\times\not{p}(...,1-n_{\xi'},...,1-n_{\xi''},...t)]\,.\end{aligned} \tag{8.25}$$

To obtain the first moment equation, we now multiply with n_{ζ^0} and sum over all $\{n_\zeta\}$. The first term in Eq. (8.25) yields, for the averaging of the factors containing n's,

$$\text{first term yields } \langle n_{\zeta^0}(1-n_{\zeta''})n_{\zeta'}\rangle_t, \tag{8.26}$$

where $\langle\ \rangle_t$ means the nonequilibrium average at t. In the second term we find

$$\text{second term yields} \begin{cases} \langle n_{\zeta^0}n_{\zeta''}(1-n_{\zeta'})\rangle_t, & \text{if } \zeta^0 \neq \zeta'', \zeta', \\ \langle (1-n_{\zeta^0})n_{\zeta''}(1-n_{\zeta'})\rangle_t, & \text{if } \zeta^0 = \zeta'' \text{ or } \zeta^0 = \zeta'. \end{cases} \tag{8.27}$$

We write Eq. (8.27) also as

$$\begin{aligned} &\langle n_{\zeta^0}n_{\zeta''}(1-n_{\zeta'})\rangle_t(1-\delta_{\zeta^0\zeta''}-\delta_{\zeta^0\zeta'}) + \langle (1-n_{\zeta^0})n_{\zeta''}(1-n_{\zeta'})\rangle_t(\delta_{\zeta^0\zeta''}+\delta_{\zeta^0\zeta'}) \\ &\quad = \langle n_{\zeta^0}n_{\zeta''}(1-n_{\zeta'})\rangle_t + \delta_{\zeta^0\zeta'}\langle n_{\zeta''}(1-n_{\zeta'})\rangle_t - \delta_{\zeta^0\zeta''}\langle n_{\zeta''}(1-n_{\zeta'})\rangle_t + 2\delta_{\zeta^0\zeta''}\langle n_{\zeta''}(1-n_{\zeta''})(1-n_{\zeta'})\rangle_t \\ &\qquad - 2\delta_{\zeta^0\zeta'}\langle n_{\zeta'}(1-n_{\zeta'})n_{\zeta''}\rangle_t. \end{aligned} \tag{8.28}$$

Since $n_\zeta = 0$ or 1, the averages in the last two terms are zero. The first term of Eq. (8.28) cancels the term (8.26) if in Eq. (8.25) we make in one of the terms the changes $\zeta' \to \zeta''$, $\zeta'' \to \zeta'$, $\eta' \to \eta''$, and $\eta'' \to \eta'$. Thus, the remaining contributions of Eq. (8.28) give

$$\begin{aligned} \sum_{\{n_\zeta\}\{N_\eta\}} n_{\zeta^0} Mp(\gamma,t) &= \sum_{\zeta''\zeta'\eta''\eta'} Q(\zeta''\eta'';\zeta'\eta')[\langle n_{\zeta''}(1-n_{\zeta'})\rangle_t\delta_{\zeta^0\zeta''} - \langle n_{\zeta''}(1-n_{\zeta'})\rangle_t\delta_{\zeta^0\zeta'}]\langle N_{\eta''}(1+N_{\eta'})\rangle_{\text{eq}} \\ &= \sum_{\zeta'\eta'\eta''} [Q(\zeta^0\eta'';\zeta'\eta')\langle n_{\zeta^0}(1-n_{\zeta'})\rangle_t\langle N_{\eta''}(1+N_{\eta'})\rangle_{\text{eq}} - Q(\zeta'\eta';\zeta^0\eta'')\langle n_{\zeta'}(1-n_{\zeta^0})\rangle_t \\ &\quad \times \langle N_{\eta'}(1+N_{\eta''})\rangle_{\text{eq}}]. \end{aligned} \tag{8.29}$$

[The two Q's in this expression are equal; see Eq. (8.18).] We also introduce the scattering kernel

$$w_{\zeta^0\zeta'} = \sum_{\eta'\eta''} Q(\zeta^0\eta'';\zeta'\eta')\langle N_{\eta''}(1+N_{\eta'})\rangle_{\text{eq}}. \tag{8.30}$$

Then

$$\begin{aligned} &\sum_{\{n_\zeta\}\{N_\eta\}} n_{\zeta^0} Mp(\gamma,t) \\ &= \sum_{\zeta'} [w_{\zeta^0\zeta'}\langle n_{\zeta^0}(1-n_{\zeta'})\rangle_t - w_{\zeta'\zeta^0}\langle n_{\zeta'}(1-n_{\zeta^0})\rangle_t]. \\ &\equiv \langle \tilde{M}_{\zeta^0} n_{\zeta^0}\rangle_t \equiv -\left(\frac{\partial\langle n_{\zeta^0}\rangle t}{\partial t}\right)_{\text{coll}}, \end{aligned} \tag{8.31}$$

where $\tilde{M}$ denotes the collision operator. Equation (8.31) is the collision "integral" of the Boltzmann equation, providing we still write, in accord with Eq. (8.1),

$$\begin{aligned} \langle n_{\zeta^0}(1-n_{\zeta'})\rangle_t &\simeq \langle n_{\zeta^0}\rangle_t(1-\langle n_{\zeta'}\rangle_t), \\ \langle n_{\zeta'}(1-n_{\zeta^0})\rangle_t &\simeq \langle n_{\zeta'}\rangle_t(1-\langle n_{\zeta^0}\rangle_t). \end{aligned} \tag{8.32}$$

We do the same for the N's in Eq. (8.30). Using for $\langle N\rangle_{\text{eq}}$ the Bose–Einstein distribution, we find for the scattering kernels the rule

$$w_{\zeta^0\zeta'} = e^{\beta(E_{\zeta^0}-E_{\zeta'})} w_{\zeta'\zeta^0}. \tag{8.33}$$

We must now find the first moment results for the other terms of the master equation (4.31). The time derivative gives the simple result

$$\sum_{\{n_\zeta\}\{N_\eta\}} n_{\zeta^0}\frac{\partial}{\partial t}p(\{n_\zeta\},\{N_\eta\},t) = \frac{\partial}{\partial t}\langle n_{\zeta^0}\rangle_t. \tag{8.34}$$

The inhomogeneous terms are evaluated by writing $(\dot{A}^R)_d$ in the second quantization form

$$(\dot{A}^R)_d = -\sum_\zeta \Lambda_d \mathbf{n}_\zeta(\zeta|a|\zeta) + \sum_\zeta \mathbf{n}_\zeta(\zeta|\dot{a}|\zeta), \tag{8.35}$$

where we used (4.28); $\mathbf{n}_\zeta = c_\zeta^+ c_\zeta$ and a and $\dot{a}$ are one-particle operators of the Schrödinger picture (we assume that the external field does not affect the bosons). For the operator $\Lambda_d \mathbf{n}_\zeta$ one may prove, by methods similar to those we used for the collision terms,

$$\langle \Lambda_d \mathbf{n}_{\zeta'}\rangle_b = \sum_{\{n_\zeta\}} |\{n_\zeta\}\rangle\langle\{n_\zeta\}|\ \tilde{M}_{\zeta'} n_{\zeta'}, \tag{8.36}$$

where $\langle\ \rangle_b$ denotes an average over the boson states, and where $\tilde{M}$ is the Boltzmann collision operator as in (8.31).

For the two streaming terms we have

$$\begin{aligned} &\beta F(t)\sum_\gamma p_{\text{eq}}(\gamma) n_{\zeta^0}\sum_{\zeta'} n_{\zeta'}(\zeta'|\dot{a}|\zeta') \\ &\quad - \beta F(t)\sum_\gamma p_{\text{eq}}(\gamma) n_{\zeta^0}\sum_{\zeta'}\langle\gamma|\Lambda_d \mathbf{n}_{\zeta'}|\gamma\rangle(\zeta'|a|\zeta'). \end{aligned} \tag{8.37}$$

For the first term we find

$$\text{first streaming term} = \beta F(t)\sum_{\zeta'}\langle n_{\zeta^0}n_{\zeta'}\rangle_{\text{eq}}(\zeta'|\dot{a}|\zeta'), \tag{8.38}$$

where the subscript "eq" denotes an average in a grand canonical ensemble. To further reduce this we write, using (8.1),

$$\begin{aligned} &\sum_{\zeta'}\langle n_{\zeta^0}n_{\zeta'}\rangle_{\text{eq}}(\zeta'|\dot{a}|\zeta') \\ &= \sum_{\zeta'\neq\zeta^0}\langle n_{\zeta^0}n_{\zeta'}\rangle_{\text{eq}}(\zeta'|\dot{a}|\zeta') + \langle n_{\zeta^0}^2\rangle_{\text{eq}}(\zeta^0|\dot{a}|\zeta^0) \\ &\approx \sum_{\zeta'\neq\zeta^0}\langle n_{\zeta^0}\rangle_{\text{eq}}\langle n_{\zeta'}\rangle_{\text{eq}}(\zeta'|\dot{a}|\zeta') + \langle n_{\zeta^0}^2\rangle_{\text{eq}}(\zeta^0|\dot{a}|\zeta^0). \end{aligned} \tag{8.39}$$

We now use (6.22′) which leads to

$$\langle \dot{A}\rangle_{\text{eq}} = \sum_{\zeta'}\langle \mathbf{n}_{\zeta'}\rangle_{\text{eq}}(\zeta'|\dot{a}|\zeta') = 0 \tag{8.40}$$

so that

$$\sum_{\zeta' \neq \zeta''} \langle n_{\zeta'} \rangle_{eq} (\zeta' | \dot{a} | \zeta') = - \langle n_{\zeta''} \rangle_{eq} (\zeta^0 | \dot{a} | \zeta^0). \tag{8.40a}$$

Equation (8.40a) into (8.39) yields

$$\sum_{\zeta'} \langle n_{\zeta''} n_{\zeta'} \rangle_{eq} (\zeta' | \dot{a} | \zeta') = [\langle n^2_{\zeta''} \rangle_{eq} - \langle n_{\zeta''} \rangle^2_{eq}] (\zeta^0 | \dot{a} | \zeta^0). \tag{8.40b}$$

Now, according to statistical mechanics,

$$\langle n^2_{\zeta''} \rangle_{eq} - \langle n_{\zeta''} \rangle^2_{eq} = \langle \Delta n^2_{\zeta''} \rangle_{eq} = \langle n_{\zeta''} \rangle_{eq} [1 - \langle n_{\zeta''} \rangle_{eq}]. \tag{8.40c}$$

The field term (8.38) thus becomes

$$\text{first field term} = \beta F(t) \langle n_{\zeta''} \rangle_{eq} [1 - \langle n_{\zeta''} \rangle_{eq}] (\zeta^0 | \dot{a} | \zeta^0). \tag{8.41}$$

The second field term is computed by using (8.36). Thus

$$\langle \gamma | \langle \Lambda_d n_{\zeta'} \rangle_b | \gamma \rangle = \tilde{M}_{\zeta' \zeta''} n_{\zeta''} \tag{8.42}$$

and the term becomes, affecting a further average over the fermion states,

$$\text{second field term} = -\beta F(t) \sum_{\zeta'} \langle n_{\zeta''} \tilde{M}_{\zeta' \zeta''} n_{\zeta'} \rangle_{eq} (\zeta' | a | \zeta'). \tag{8.43}$$

This result can be simplified as follows. We write

$$\sum_{\zeta'} \langle n_{\zeta''} \tilde{M} n_{\zeta'} \rangle_{eq} (\zeta' | a | \zeta') = \sum_{\zeta' \neq \zeta^0, \zeta''} (\zeta' | a | \zeta') [w_{\zeta' \zeta''} \langle n_{\zeta''} n_{\zeta'} (1 - n_{\zeta''}) \rangle_{eq} - w_{\zeta'' \zeta'} \langle n_{\zeta''} n_{\zeta''} (1 - n_{\zeta'}) \rangle_{eq}] \tag{1}$$

$$+ \sum_{\zeta''} (\zeta^0 | a | \zeta^0) [w_{\zeta^0 \zeta''} \langle n^2_{\zeta''} (1 - n_{\zeta''}) \rangle_{eq} - w_{\zeta'' \zeta^0} \langle n_{\zeta''} n_{\zeta''} (1 - n_{\zeta^0}) \rangle_{eq}]. \tag{2}$$

Term (1) we split in two parts and we use (8.1):

$$(1) = \sum_{\zeta' \neq \zeta^0} (\zeta' | a | \zeta') \sum_{\zeta'' \neq \zeta^0} [w_{\zeta' \zeta''} \langle n_{\zeta''} \rangle_{eq} \langle n_{\zeta'} \rangle_{eq} (1 - \langle n_{\zeta''} \rangle_{eq}) - w_{\zeta'' \zeta'} \langle n_{\zeta''} \rangle_{eq} \langle n_{\zeta''} \rangle_{eq} (1 - \langle n_{\zeta'} \rangle_{eq})] \tag{1a}$$

$$+ \sum_{\zeta' \neq \zeta^0} (\zeta' | a | \zeta') [w_{\zeta' \zeta^0} \langle n_{\zeta^0} (1 - n_{\zeta^0}) \rangle_{eq} \langle n_{\zeta'} \rangle_{eq} - w_{\zeta^0 \zeta'} \langle n^2_{\zeta^0} \rangle_{eq} (1 - \langle n_{\zeta'} \rangle_{eq})]. \tag{1b}$$

Using detailed balancing,

$$w_{\zeta'' \zeta'} \langle n_{\zeta''} \rangle_{eq} (1 - \langle n_{\zeta'} \rangle_{eq}) = w_{\zeta' \zeta''} \langle n_{\zeta'} \rangle_{eq} (1 - \langle n_{\zeta''} \rangle_{eq}),$$

we see that term (1a) is zero. The restriction $\zeta' \neq \zeta^0$ in term (1b) can be omitted since $w_{\zeta' \zeta^0} = 0$ for $\zeta' = \zeta^0$. Term (2) is written as

$$(2) = - \sum_{\zeta'} (\zeta^0 | a | \zeta^0) [w_{\zeta' \zeta^0} \langle n^0_{\zeta^0} (1 - n_{\zeta^0}) \rangle_{eq} \langle n_{\zeta'} \rangle_{eq} - w_{\zeta^0 \zeta'} \langle n^2_{\zeta^0} \rangle_{eq} (1 - \langle n_{\zeta'} \rangle_{eq})]. \tag{2a}$$

We now combine (1b) and (2a) to read

$$\sum_{\zeta'} [(\zeta' | a | \zeta') - (\zeta^0 | a | \zeta^0)] [w_{\zeta' \zeta^0} \langle n_{\zeta^0} (1 - n_{\zeta^0}) \rangle_{eq} \langle n_{\zeta'} \rangle_{eq} - w_{\zeta^0 \zeta'} \langle n^2_{\zeta^0} \rangle_{eq} (1 - \langle n_{\zeta'} \rangle_{eq})]. \tag{3}$$

In these terms we write

$$\langle n_{\zeta''} (1 - n_{\zeta''}) \rangle_{eq} = \langle n_{\zeta''} \rangle_{eq} (1 - \langle n_{\zeta''} \rangle_{eq}) - \langle \Delta n^2_{\zeta''} \rangle_{eq}, \tag{8.44}$$

$$\langle n^2_{\zeta''} \rangle_{eq} = \langle \Delta n^2_{\zeta''} \rangle_{eq} + \langle n_{\zeta''} \rangle^2_{eq}.$$

We then arrive at

$$(3) = \sum_{\zeta'} [(\zeta' | a | \zeta') - (\zeta^0 | a | \zeta^0)] \times [w_{\zeta' \zeta''} \langle n_{\zeta''} \rangle_{eq} (1 - \langle n_{\zeta''} \rangle_{eq} \langle n_{\zeta'} \rangle_{eq} - w_{\zeta^0 \zeta'} \langle n_{\zeta^0} \rangle^2_{eq} (1 - \langle n_{\zeta'} \rangle_{eq})] \tag{3a}$$

$$+ \sum_{\zeta'} [(\zeta^0 | a | \zeta^0) - (\zeta' | a | \zeta')] \times [w_{\zeta^0 \zeta'} \langle \Delta n^2_{\zeta''} \rangle_{eq} (1 - \langle n_{\zeta'} \rangle_{eq}) + w_{\zeta' \zeta^0} \langle \Delta n^2_{\zeta''} \rangle \langle n_{\zeta'} \rangle_{eq}]. \tag{3b}$$

From detailed balancing, (3a) is seen to be zero. Thus the result is (3b). So finally the second field term (8.43) becomes, using also (8.40c),

$$\text{second field term} = -\beta F(t) \langle n_{\zeta''} \rangle_{eq} (1 - \langle n_{\zeta''} \rangle_{eq}) \times \sum_{\zeta'} [(\zeta^0 | a | \zeta^0) - (\zeta' | a | \zeta')] \times [w_{\zeta^0 \zeta'} (1 - \langle n_{\zeta'} \rangle_{eq}) + \omega_{\zeta' \zeta^0} \langle n_{\zeta'} \rangle_{eq}]. \tag{8.45}$$

We now collect all terms. The full quantum mechanical Boltzmann equation becomes, dropping the superscript zero on ζ^0,

$$\frac{\partial}{\partial t} \langle n_\zeta \rangle_t - \beta F(t) \langle n_\zeta \rangle_{eq} (1 - \langle n_\zeta \rangle_{eq}) \times \left\{ (\zeta | \dot{a} | \zeta) + \sum_{\zeta'} [(\zeta' | a | \zeta') - (\zeta | a | \zeta)] \times [w_{\zeta \zeta'} (1 - \langle n_{\zeta'} \rangle_{eq}) + w_{\zeta' \zeta} \langle n_{\zeta'} \rangle_{eq}] \right\}$$

$$= \sum_{\zeta'} [w_{\zeta' \zeta} \langle n_{\zeta'} \rangle_t (1 - \langle n_\zeta \rangle_t) - w_{\zeta \zeta'} \langle n_\zeta \rangle_t (1 - \langle n_{\zeta'} \rangle_t)]. \tag{8.46}$$

The above is easily carried over for the case that the particles for which the transport is needed are bosons. Repeating verbatim the derivations, one finds that (8.46) only requires the changes

$$1 - \langle n_\zeta \rangle \to 1 + \langle n_\zeta \rangle \tag{8.47}$$

wherever it occurs.

We remark that there are two streaming terms, (8.41) and (8.43) or (8.45), which correspond to the one classical term $(\mathbf{F}/\hbar) \cdot \nabla_{\mathbf{k}} n_{\mathbf{k}}$ (the gradient term involving $\mathbf{v}_{\mathbf{k}} \cdot \nabla_{\mathbf{r}} n$ does not occur in a strict quantum treatment; it requires the introduction of the Wigner function). The term (8.45) indicates streaming as affected by collisions. Thus, the present Boltzmann equation carries us beyond the first Chapman–Enskog approximation. In applications we will see that one or the other field term occurs. For ordinary conductivity without a magnetic field only the first field term has nonzero matrix elements.

When we now employ the new Boltzmann equation to

calculate a current $(\dot{B}^R)_d$, we must realize that there are again two contributions, analogous to (8.35); using (8.36), we have

$$\langle J_{B,d}\rangle = \langle(\dot{B}^R)_d\rangle$$
$$= \sum_{\{n_\zeta\}} p_{eq}(\{n_\zeta\}) \sum_\zeta \times [-\tilde{M}_\zeta n_\zeta(\zeta|a|\zeta) + n_\zeta(\zeta|\dot{a}|\zeta)]$$
$$= \sum_\zeta \left\{ \left(\frac{\partial \langle n_\zeta\rangle}{\partial t} \right)_{coll} (\zeta|a|\zeta) + \langle n_\zeta\rangle(\zeta|\dot{a}|\zeta) \right\}. \tag{8.47a}$$

As noted in Sec. 4, the two terms of (8.47a) represent the collisional current and the ponderomotive current. Only the latter reflects the motion in the subdynamics of $\mathscr{H}^0$. For that reason the first term has usually been omitted, though it appeared in a hidden fashion in some treatments.[18] (We note, in this respect, that in the *full* dynamics of $\mathscr{H}$ there is only ponderomotive current, $J = \Sigma_i v_i$.) Since the Boltzmann equation pertains to the subdynamics of $\mathscr{H}^0$, the collisional current cannot be overlooked. In fact, we will show in a future paper[20] that it is the main current in transverse magneto resistance phenomena, since the matrix elements $(\zeta|\dot{x}|\zeta)$ and $(\zeta|\dot{y}|\zeta)$ vanish for Landau states.

We illustrate this result for the **k**-space formulation of electrons in solids subject to an electrical field. Then the field Hamiltonian is

$$-AF(t) = e\mathbf{E}\cdot \sum_i (\mathbf{q}_i - \langle \mathbf{q}_{i0}\rangle_{eq}), \tag{8.48}$$

where $\mathbf{q}_i$ are the electron positions. and $\mathbf{q}_{i0}$ the positions prior to application of the field.Thus, $F = -e\mathbf{E}\cdot$ and $a = \mathbf{q} - \langle \mathbf{q}_0\rangle_{eq}$. The one-particle matrix elements are, if $|\zeta) \equiv |\mathbf{k}) = \exp(i\mathbf{k}\cdot\mathbf{q})u_\mathbf{k}(\mathbf{q})$ is a Bloch state,

$$(\mathbf{k}|\mathbf{q} - \langle \mathbf{q}_0\rangle_{eq}|\mathbf{k}) = \langle \mathbf{q}\rangle_\mathbf{k} - \langle \mathbf{q}_0\rangle_{eq} = 0,$$
$$(\mathbf{k}|\dot{\mathbf{q}}|\mathbf{k}) = \hbar^{-1}\nabla_\mathbf{k} e(\mathbf{k}), \tag{8.49}$$

where $e(\mathbf{k})$ is the Bloch state energy. Employing for $\langle n_\mathbf{k}\rangle_{eq}$ Fermi–Dirac statistics, the streaming term, Eq. (8.41) becomes

$$-\beta \frac{e\mathbf{E}}{\hbar}\cdot \frac{1}{e^{\beta[e(\mathbf{k}) - e_f]} + 1} \frac{1}{e^{-\beta[e(\mathbf{k}) - e_f]} + 1} \nabla_\mathbf{k} e(\mathbf{k})$$
$$= \frac{e\mathbf{E}}{\hbar}\cdot \nabla_\mathbf{k} \left[\frac{1}{e^{\beta[e(\mathbf{k}) - e_f]} + 1} \right], \tag{8.50}$$

which is the standard result for the streaming term in the perturbation form of the Boltzmann equation.

For the **k**-space formalism the states are dense and we can introduce the coarse grained density

$$f(\mathbf{k},t) = \langle n_\mathbf{k}\rangle_t Z(\mathbf{k}), \quad \int f(\mathbf{k}) d^3k = \mathscr{N},$$

where $Z(\mathbf{k})$ is the density of states defined by

$$\sum_\mathbf{k} = \int Z(\mathbf{k}) d^3k = \frac{\Omega}{4\pi^3} \int d^3k,$$

and where $\mathscr{N}$ is the total number of fermions. Then Eq. (8.46) takes the usual form

$$\frac{\partial}{\partial t} f(\mathbf{k},t) - (e\mathbf{E}/\hbar)\nabla_\mathbf{k} f_{eq}(\mathbf{k})$$
$$= \int z(\mathbf{k}) d^3k' \{ w_{\mathbf{k}'\mathbf{k}} f(\mathbf{k}',t)[1 - f(\mathbf{k},t)/Z(\mathbf{k})]$$
$$- w_{\mathbf{k}\mathbf{k}'} f(\mathbf{k},t)[1 - f(\mathbf{k}',t)/Z(\mathbf{k}')] \}. \tag{8.51}$$

B. Electron–phonon interaction

For an electron–phonon system the interaction Hamiltonian is[21]

$$\lambda V = i \sum_{\mathbf{k}'\mathbf{q}'} F(\mathbf{q}')(a_{\mathbf{q}'} c^\dagger_{\mathbf{k}'+\mathbf{q}'} c_{\mathbf{k}'} - a^\dagger_{\mathbf{q}'} c^\dagger_{\mathbf{k}'-\mathbf{q}'} c_{\mathbf{k}'}), \tag{8.52}$$

where $|\mathbf{k})$ is an electron Bloch state, $|\mathbf{q})$ is a phonon state, $c, c^\dagger$ and $a, a^\dagger$ are the annihilation and creation operators for electrons and phonons, respectively, and where $F(\mathbf{q})$ is an interaction function depending on the model (deformation potential model, rigid ion approximation model, etc.). The states $|\gamma\rangle$ are denoted as $|\{n_\mathbf{k}\},\{N_\mathbf{q}\}\rangle$. The raising and lowering rules are as in Eqs. (8.10)–(8.12). We divide λV in two parts λV_1 and λV_2, where the first refers to the first term in Eq. (8.52), which is associated with phonon absorption, whereas the second refers to the second term in Eq. (8.52), which represents phonon emission. We find that matrix elements $\langle\bar{\gamma}|\lambda V_1|\gamma\rangle$ exist between states such that, for given $\mathbf{k}'$ and $\mathbf{q}'$ of the summand of Eq. (8.52), we have

$$\bar{n}_\mathbf{k} = n_\mathbf{k}, \quad \text{for } \mathbf{k}\neq\mathbf{k}'+\mathbf{q}' \text{ and } \mathbf{k}\neq\mathbf{k}',$$
$$\bar{n}_{\mathbf{k}'} = 1 - n_{\mathbf{k}'} \text{ and } \bar{n}_{\mathbf{k}'+\mathbf{q}'} = 1 - n_{\mathbf{k}'+\mathbf{q}'},$$
$$\bar{N}_\mathbf{q} = N_\mathbf{q} - \delta_{\mathbf{q}\mathbf{q}'}. \tag{8.53}$$

The value of the matrix element in that case is

$$\langle\bar{\gamma}_{\mathbf{k}',\mathbf{q}'}|\lambda V_1|\gamma\rangle = (-1)^{\sum_\mathbf{k}} (-1)^{\sum_{\mathbf{k}<\mathbf{q}'}} iF(\mathbf{q}') \times [(1 - n_{\mathbf{k}'+\mathbf{q}'}) n_{\mathbf{k}'} N_{\mathbf{q}'}]^{1/2}. \tag{8.54}$$

Likewise, matrix elements $\langle\bar{\gamma}|\lambda V_2|\gamma\rangle$ exist for states $|\bar{\gamma}_{\mathbf{k}',\mathbf{q}'}\rangle$ connected to $|\gamma\rangle$ by

$$\bar{n}_\mathbf{k} = n_\mathbf{k}, \quad \text{for } \mathbf{k}\neq\mathbf{k}'-\mathbf{q}' \text{ and } \mathbf{k}\neq\mathbf{k}',$$
$$\bar{n}_{\mathbf{k}'} = 1 - n_{\mathbf{k}'} \text{ and } \bar{n}_{\mathbf{k}'-\mathbf{q}'} = 1 - n_{\mathbf{k}'-\mathbf{q}'},$$
$$\bar{N}_\mathbf{q} = N_\mathbf{q} + \delta_{\mathbf{q}\mathbf{q}'}. \tag{8.55}$$

The value of the matrix element in that case is

$$\langle\bar{\gamma}_{\mathbf{k}',\mathbf{q}'}|\lambda V_2|\gamma\rangle = (-1)^{\sum_\mathbf{k}} (-1)^{\sum_{\mathbf{k}<\mathbf{q}'}} iF(\mathbf{q}') \times [(1 - n_{\mathbf{k}'-\mathbf{q}'}) n_{\mathbf{k}'} (1 + N_{\mathbf{q}'})]^{1/2}. \tag{8.56}$$

We introduce now

$$Q(\mathbf{k},\mathbf{q}\to\mathbf{k}') = (2\pi/\hbar)|F(\mathbf{q})|^2 \delta(e_\mathbf{k} + E_\mathbf{q} - e_{\mathbf{k}'}) \delta_{\mathbf{k}',\mathbf{k}+\mathbf{q}} \tag{8.57a}$$

and

$$Q(\mathbf{k}\to\mathbf{k}',\mathbf{q}) = (2\pi/\hbar)|F(\mathbf{q})|^2 \delta(e_\mathbf{k} - e_{\mathbf{k}'} - E_\mathbf{q}) \delta_{\mathbf{k}',\mathbf{k}-\mathbf{q}}. \tag{8.57b}$$

The Q's are reversible:

$$Q(\mathbf{k},\mathbf{q}\to\mathbf{k}') = Q(\mathbf{k}'\to\mathbf{k},\mathbf{q}). \tag{8.58}$$

For the transition rates we easily find

$$W^{abs}_{\gamma,\gamma_{\mathbf{k}'\mathbf{q}'}} = Q(\mathbf{k}',\mathbf{q}'\to\mathbf{k}'+\mathbf{q}')(1 - n_{\mathbf{k}'+\mathbf{q}'}) n_{\mathbf{k}'} N_{\mathbf{q}'}, \tag{8.59}$$
$$W^{em}_{\gamma,\gamma_{\mathbf{k}'\mathbf{q}}} = Q(\mathbf{k}'\to\mathbf{k}'-\mathbf{q}',\mathbf{q}')(1 - n_{\mathbf{k}'-\mathbf{q}'}) n_{\mathbf{k}'} (1 + N_{\mathbf{q}'}), \tag{8.60}$$

For the master operator expression $Mp(\gamma,t)$, we now obtain

$$Mp(\{n_{\mathbf{k}}\},\{N_{\mathbf{q}}\},t) = \sum_{\mathbf{k}'\mathbf{q}'} \{Q(\mathbf{k}',\mathbf{q}'\to\mathbf{k}'+\mathbf{q}')(1-n_{\mathbf{k}'+\mathbf{q}'})n_{\mathbf{k}'}N_{\mathbf{q}'}[p(\{n_{\mathbf{k}}\},\{N_{\mathbf{q}}\},t) - p(\ldots,1-n_{\mathbf{k}'},\ldots,1-n_{\mathbf{k}'+\mathbf{q}'},\ldots,N_{\mathbf{q}'}-1,\ldots,t) + Q(\mathbf{k}'\to\mathbf{k}'-\mathbf{q}',\mathbf{q}')(1-n_{\mathbf{k}'-\mathbf{q}'})n_{\mathbf{k}'}(1+N_{\mathbf{q}'})[p(\{n_{\mathbf{k}}\},\{N_{\mathbf{q}}\},t)] - p(\ldots,1-n_{\mathbf{k}'-\mathbf{q}'},\ldots,1-n_{\mathbf{k}'},\ldots,N_{\mathbf{q}'}+1,\ldots,t)]\}. \tag{8.61}$$

Summing over the phonon states, this leads to

$$\sum_{\{N_{\mathbf{q}}\}} Mp\{n_{\mathbf{k}}\},\{N_{\mathbf{q}}\},t) = \sum_{\mathbf{k}'\mathbf{q}'} \{Q(\mathbf{k}',\mathbf{q}'\to\mathbf{k}'+\mathbf{q}')(1-n_{\mathbf{k}'+\mathbf{q}'})n_{\mathbf{k}'}[\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}}\not{e}(\{n_{\mathbf{k}}\},t) - (1+\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}})\not{e}(\ldots,1-n_{\mathbf{k}'},\ldots,1-n_{\mathbf{k}'+\mathbf{q}'},\ldots,t] + Q(\mathbf{k}'\to\mathbf{k}'-\mathbf{q}',\mathbf{q}')(1-n_{\mathbf{k}'-\mathbf{q}'})n_{\mathbf{k}'}[(1+\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}})\not{e}(\{n_{\mathbf{k}}\},t) - \langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}}\not{e}(\ldots,1-n_{\mathbf{k}'-\mathbf{q}'},\ldots,1-n_{\mathbf{k}'},\ldots,t)]\}. \tag{8.62}$$

We now multiply with $n_{\mathbf{k}}$ and sum over $\{n_{\mathbf{k}}\}$. This summation proceeds as in the previous subsection. We find

$$\sum_{\{n_{\mathbf{k}}\}\{N_{\mathbf{q}}\}} n_{\mathbf{k}}Mp(\{n_{\mathbf{k}}\},\{N_{\mathbf{q}}\},t) = \sum_{\mathbf{k}',\mathbf{q}'} \{Q(\mathbf{k}',\mathbf{q}'\to\mathbf{k}'+\mathbf{q}')\langle(1-n_{\mathbf{k}'+\mathbf{q}'})n_{\mathbf{k}'}n_{\mathbf{k}}\rangle_t\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}} - Q(\mathbf{k}',\mathbf{q}'\to\mathbf{k}'+\mathbf{q}')[\langle(1-n_{\mathbf{k}'})n_{\mathbf{k}'+\mathbf{q}'}n_{\mathbf{k}}\rangle_t - \delta_{\mathbf{k},\mathbf{k}'+\mathbf{q}'}\langle n_{\mathbf{k}'+\mathbf{q}'}(1-n_{\mathbf{k}'})\rangle_t + \delta_{\mathbf{k}\mathbf{k}'}\langle n_{\mathbf{k}'+\mathbf{q}'}(1-n_{\mathbf{k}'}\rangle_t](1+\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}}) + Q(\mathbf{k}'\to\mathbf{k}'-\mathbf{q}',\mathbf{q}')\langle(1-n_{\mathbf{k}'-\mathbf{q}'})n_{\mathbf{k}'}n_{\mathbf{k}}\rangle_t(1+\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}}) - Q(\mathbf{k}'\to\mathbf{k}'-\mathbf{q}',\mathbf{q}')[\langle(1-n_{\mathbf{k}'})n_{\mathbf{k}'-\mathbf{q}'}n_{\mathbf{k}}\rangle_t - \delta_{\mathbf{k},\mathbf{k}'-\mathbf{q}'}\langle n_{\mathbf{k}'-\mathbf{q}'}(1-n_{\mathbf{k}'})\rangle_t + \delta_{\mathbf{k}\mathbf{k}'}\langle n_{\mathbf{k}'-\mathbf{q}'}(1-n_{\mathbf{k}'})\rangle_t]\langle N_{\mathbf{q}'}\rangle_{\mathrm{eq}}.$$

There are four terms involving a Q factor and a triple correlation $\langle(1-n)nn\rangle_t$. In the first term we substitute $\mathbf{k}'\to\mathbf{k}'-\mathbf{q}'$. It then cancels the fourth. In the third term we substitute $\mathbf{k}'\to\mathbf{k}'+\mathbf{q}'$; it then cancels the second. The remaining terms then yield, observing Eqs. (8.58) and (8.1),

$$\sum_{\{n_{\mathbf{k}}\}\{N_{\mathbf{q}}\}} n_{\mathbf{k}}Mp(\{n_{\mathbf{k}}\},\{N_{\mathbf{q}}\},t) = -\left(\frac{\partial}{\partial t}\langle n_{\mathbf{k}}\rangle_t\right)_{\mathrm{coll}} = \sum_{\mathbf{q}} Q(\mathbf{k},\mathbf{q}\to\mathbf{k}+\mathbf{q})\langle n_{\mathbf{k}}\rangle_t(1-\langle n_{\mathbf{k}+\mathbf{q}}\rangle_t)\langle N_{\mathbf{q}}\rangle_{\mathrm{eq}} - Q(\mathbf{k}-\mathbf{q},\mathbf{q}\to\mathbf{k})\langle n_{\mathbf{k}-\mathbf{q}}\rangle_t(1-\langle n_{\mathbf{k}}\rangle_t)\langle N_{\mathbf{q}}\rangle_{\mathrm{eq}} + Q(\mathbf{k}\to\mathbf{k}-\mathbf{q},\mathbf{q})\langle n_{\mathbf{k}}\rangle_t(1-\langle n_{\mathbf{k}-\mathbf{q}}\rangle_t)(1+\langle N_{\mathbf{q}}\rangle_{\mathrm{eq}}) - Q(\mathbf{k}+\mathbf{q}\to\mathbf{k},\mathbf{q})\langle n_{\mathbf{k}+\mathbf{q}}\rangle_t(1-\langle n_{\mathbf{k}}\rangle_t)(1+N_{\mathbf{q}}\rangle_{\mathrm{eq}}). \tag{8.63}$$

The four terms correspond to gain or loss due to photon absorption or emission. Introducing

$$w_{\mathbf{k}\mathbf{k}'} = \sum_{\mathbf{q}} [Q(\mathbf{k},\mathbf{q}\to\mathbf{k}')\langle N_{\mathbf{q}}\rangle_{\mathrm{eq}} + Q(\mathbf{k}\to\mathbf{k}',\mathbf{q})(1+\langle N_{\mathbf{q}}\rangle_{\mathrm{eq}})], \tag{8.64}$$

we find that Eq. (8.63) takes the same form as in the previous subsection—noting the Kronecker deltas in the definitions of the Q's [Eqs. (8.57) and (8.58)],

$$\left(\frac{\partial}{\partial t}\langle n_{\mathbf{k}}\rangle_t\right)_{\mathrm{coll}} = \sum_{\mathbf{k}'} [w_{\mathbf{k}'\mathbf{k}}\langle n_{\mathbf{k}'}\rangle_t(1-\langle n_{\mathbf{k}}\rangle_t) - w_{\mathbf{k}\mathbf{k}'}\langle n_{\mathbf{k}}\rangle_t(1-\langle n_{\mathbf{k}'}\rangle_t)]. \tag{8.65}$$

The field terms are identical to that of the previous subsection. Thus, the final Boltzmann equation is again Eq. (8.46). A proof of Boltzmann's H- theorem for the case of electron–phonon collisions was given recently.[22]

ACKNOWLEDGMENTS

I would like to thank Professor L. Garcia-Colin for his critical reading of a previous version of the manuscript and for discussions on the problem of irreversibility. I also thank an unknown reviewer, whose criticism led us to a sharper formulation of the projector algebra problems mentioned in Sec. 2 and Appendix B. I am indebted to M. Charbonneau for preliminary results on the transverse magneto resistance, which led to a more complete form of the Boltzmann equation. Discussions with L.de Sobrino, N.G.van Kampen, and M. Grmela have been highly appreciated.

APPENDIX A: ON TETRADIC REPRESENTATION

From the first equality of Eq. (2.1) it follows that the identity superoperator allows the symbolic notation

$$\mathscr{I} = \sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma|\overrightarrow{\cdot}\overleftarrow{\cdot}|\gamma'\rangle\langle\gamma'|, \tag{A1}$$

where the arrows indicate that the first operator works on K to the right and the second operator works on K to the left. A similar notation holds for the Liouville operator

$$\mathscr{L} = \frac{1}{\hbar}[H\overrightarrow{\cdot}I - I\overleftarrow{\cdot}H], \tag{A2}$$

where I is the unit operator of the state space which contains the dynamical variables. Clearly, the Liouville space is of dimension $\mathscr{H}\otimes\overline{\mathscr{H}}$ (cf. Fano[7]). Whereas ordinary operators have a dyadic representation $K_{\alpha\beta} = \langle\alpha|K|\beta\rangle$, the superoperators allow a tetradic representation. The four subscripts are arranged as follows:

$$\mathscr{I}_{\alpha\beta|\alpha'\beta'} = \sum_{\gamma\gamma'}\langle\alpha|\{|\gamma\rangle\langle\gamma|\}|\alpha'\rangle\langle\beta'|\{|\gamma'\rangle\langle\gamma'|\}|\beta\rangle = \sum_{\gamma\gamma'}\langle\alpha|\gamma\rangle\langle\gamma|\alpha'\rangle\langle\beta'|\gamma'\rangle\langle\gamma'|\beta'\rangle = \delta_{\alpha\alpha'}\delta_{\beta\beta'}. \tag{A3}$$

Likewise,

$$\begin{aligned}\mathscr{L}_{\alpha\beta|\alpha'\beta'} &= (1/\hbar)(\langle\alpha|H|\alpha'\rangle\langle\beta'|I|\beta\rangle \\ &\quad - \langle\alpha|I|\alpha'\rangle\langle\beta'|H|\beta\rangle) \\ &= (1/\hbar)(\delta_{\beta\beta'}\langle\alpha|H|\alpha'\rangle - \delta_{\alpha\alpha'}\langle\beta'|H|\beta\rangle) \\ &= (1/\hbar)(\delta_{\beta\beta'}H_{\alpha\alpha'} - \delta_{\alpha\alpha'}H_{\beta'\beta}). \end{aligned} \tag{A4}$$

The superoperator rule is

$$(\mathscr{S}K)_{\alpha\beta} = \sum_{\alpha'\beta'} \mathscr{S}_{\alpha\beta|\alpha'\beta'}K_{\alpha'\beta'}. \tag{A5}$$

For the identity operator we easily check that

$$(\mathscr{I}K)_{\alpha\beta} = \sum_{\alpha'\beta'} \delta_{\alpha\alpha'}\delta_{\beta\beta'}K_{\alpha'\beta'} = K_{\alpha\beta}, \tag{A6}$$

so that $\mathscr{I}K = K$. Likewise for $\mathscr{L}$ we verify

$$\begin{aligned}(\mathscr{L}K)_{\alpha\beta} &= \sum_{\alpha'\beta'} \mathscr{L}_{\alpha\beta|\alpha'\beta'}K_{\alpha'\beta'} \\ &= \sum_{\alpha'\beta'} \frac{1}{\hbar}(\delta_{\beta\beta'}H_{\alpha\alpha'} - \delta_{\alpha\alpha'}H_{\beta'\beta})K_{\alpha'\beta'} \\ &= (1/\hbar)[(HK)_{\alpha\beta} - (KH)_{\alpha\beta}], \end{aligned} \tag{A7}$$

so that $\mathscr{L}K = (1/\hbar)[H,K]$ in accord with Eq. (2.9).

Finally, for the projection operator we have the notation

$$\mathscr{P} = \sum_{\gamma} |\gamma\rangle\langle\gamma|^{\rightarrow\leftarrow}|\gamma\rangle\langle\gamma| \tag{A8}$$

and for Λ_d we have

$$\begin{aligned}\Lambda_d = -\sum_{\gamma\gamma''} [&W_{\gamma''\gamma}|\gamma\rangle\langle\gamma''|^{\rightarrow\leftarrow}|\gamma''\rangle\langle\gamma| \\ &- W_{\gamma\gamma''}|\gamma\rangle\langle\gamma|^{\rightarrow\leftarrow}|\gamma\rangle\langle\gamma|]. \end{aligned} \tag{A9}$$

Thus,

$$\begin{aligned}\mathscr{P}_{\alpha\beta|\alpha'\beta'} &= \sum_{\gamma} \langle\alpha|\gamma\rangle\langle\gamma|\alpha'\rangle\langle\beta'|\gamma\rangle\langle\gamma|\beta\rangle \\ &= \delta_{\alpha\alpha'}\delta_{\beta\beta'}\delta_{\alpha\beta}, \end{aligned} \tag{A10}$$

in accord with Zwanzig's tetradic representation for $\mathscr{P}$. We note that

$$(\mathscr{P}K)_{\alpha\beta} = \sum_{\alpha'\beta'} \delta_{\alpha\alpha'}\delta_{\beta\beta'}\delta_{\alpha\beta}K_{\alpha'\beta'} = \delta_{\alpha\beta}K_{\alpha\beta}, \tag{A11}$$

as expected. For Λ_d we obtain

$$\Lambda_{d\alpha\beta|\alpha'\beta'} = -\sum_{\gamma} (W_{\gamma\alpha}\delta_{\alpha\beta}\delta_{\alpha'\beta'}\delta_{\gamma\alpha'} - W_{\alpha\gamma}\delta_{\alpha\alpha'}\delta_{\beta\beta'}\delta_{\alpha\beta}). \tag{A12}$$

Whereas the above relations are at times useful to verify projector calculus rules, generally we find it preferable to work with an abstract Liouville space rather than its tetrad space realization.

APPENDIX B: ON THE DIRAC BILINEAR FORM

As shown in Sec. 2, the Dirac bilinear form $\langle\phi|K|\psi\rangle$ presents problems if the bilinear concommitant is not equal to zero. Specifically, we have

$$\begin{aligned}\langle\phi|\{K|\psi\rangle\} - \{\langle\phi|K\}|\psi\rangle &= (K\psi,\phi) - (\psi,K^{\dagger}\phi) \\ &= Q_K(\psi,\phi), \end{aligned} \tag{B1}$$

where $Q_K(\psi,\phi)$ is a surface integral. If $K = K^{\dagger}$, the operator K is Hermitian; if in addition the boundary conditions on ϕ and ψ, necessary for Q_K to vanish, are identical, K is self-adjoint. Because of the possible nonassociativity of the Dirac bilinear form, we must specify the matrix elements in a definite way. Thus, we define

$$K_{\gamma\gamma'} \equiv \langle\gamma|\{K|\gamma'\rangle\} = (K\gamma',\gamma). \tag{B2}$$

The diagonal part of K will be written as

$$\mathscr{P}K = \sum_{\gamma} R_{\gamma}(K\gamma,\gamma), \tag{B3}$$

where γ are the eigenstates and R_{γ} are the projectors of H^0. Though there is no easy notation for R_{γ} in algebric notation, the concept is equally well valid as in Dirac notation (see, for example, Riesz and Nagy[23]). It follows the rule

$$R_{\gamma}\phi = (\phi,\gamma)\gamma. \tag{B4}$$

We now consider the diagonal part of a commutator $[K,H^0]$. For the matrix element we have, using Eq. (B1),

$$\begin{aligned}([K,H^0]\gamma,\gamma) &= (KH^0\gamma,\gamma) - (H^0K\gamma,\gamma) \\ &= (H^0\gamma,K\gamma) + Q_K(H^0\gamma,\gamma) - (K\gamma,H^0\gamma) \\ &\quad - Q_{H^0}(K\gamma,\gamma) \\ &= \epsilon_{\gamma}(\gamma,K\gamma) + \epsilon_{\gamma}Q_K(\gamma,\gamma) - \epsilon_{\gamma}(K\gamma,\gamma) \\ &\quad - Q_{H^0}(K\gamma,\gamma) \\ &= -Q_{H^0}(K\gamma,\gamma). \end{aligned} \tag{B5}$$

So this only vanishes if $Q_{H^0}(K\gamma,\gamma)$ vanishes, i.e, when $K\gamma$ satisfies the same boundary conditions as γ. For the diagonal part we find

$$\mathscr{P}[K,H^0] = -\sum_{\gamma} R_{\gamma}Q_{H^0}(K\gamma,\gamma). \tag{B6}$$

We now show the correctness of these results by two examples, involving the one-particle Hamiltonian

$$H^0 = p^2/2m + U(q). \tag{B7}$$

We consider the commutator $[q,H^0]$. As is well known from the Poisson bracket relation,

$$[q,H^0] = \hbar i(\partial H^0/\partial p) = \hbar i(p/m). \tag{B8}$$

Let $U(q) = \frac{1}{2}\alpha q^2$ pertain to an harmonic oscillator. Then, for the matrix element of the commutator (B8),

$$([q,H^0]\gamma,\gamma) = (\hbar i/m)(p\gamma,\gamma) = 0, \tag{B9}$$

since for an harmonic oscillator the matrix $(p\gamma,\gamma')$ has zeros for the diagonal $\gamma = \gamma'$. Thus, $\mathscr{P}[q,H^0] = 0$ by the above direct calculation. It is also zero from Eq. (B6), for the bilinear concomitant vanishes since for the harmonic oscillator, all states, γ, as well as $K\gamma$, go sufficiently fast to zero at the boundaries $q = \pm\infty$ of the system.

Now let $U(q)$be $-U_0$ for $-L/2 \leqslant q \leqslant L/2$ (particle in a box) and let γ satisfy periodic boundary conditions. The eigenstates are $|\gamma\rangle \equiv |k\rangle$, or in wave language

$$\gamma = L^{-1/2}e^{ik_{\gamma}q},\ k_{\gamma} = 2\pi n/L,\ n = 0, \pm 1, \ldots\ . \tag{B10}$$

For the commutator matrix elements we have

$$([q,H^0]\gamma,\gamma) = (\hbar i/m)(p\gamma,\gamma) = (\hbar^2 i/m)k_{\gamma}. \tag{B11}$$

Thus, the diagonal part of the commutator does not vanish. [Moreover, we note that, for the operator p, $(p\gamma,\gamma)$

$= (\gamma,p\gamma) = \langle\gamma|p|\gamma\rangle =$ associative, so that the same result is obtained, had we defined the matrix element $K_{\gamma\gamma'}$ as $(\gamma',K\gamma)$ instead of by Eq. (B2); thus, the result is unique.] We next show that the same answer obtains from Eq. (B5). Since for the box particle $H^0 = (-\hbar^2/2m)d^2/dq^2 - U_0$, we have by the one-dimensional Green's theorem

$$Q_{H^0}(\phi,\psi) = \frac{\hbar^2}{2m}\left[\phi\frac{d\psi^*}{dq} - \psi^*\frac{d\phi}{dq}\right]_{-L/2}^{L/2}. \tag{B12}$$

Thus,

$$\begin{aligned} -Q_{H^0}(q\gamma,\gamma) = &-\frac{\hbar^2}{2mL}\left[q\,e^{ik_\gamma q}\frac{d}{dq}e^{-ik_\gamma q}\right. \\ &\left. - e^{-ik_\gamma q}\frac{d}{dq}(q\,e^{ik_\gamma q})\right]_{-L/2}^{L/2} \\ = &(\hbar^2/m)ik_\gamma, \end{aligned} \tag{B13}$$

which is the same as the result by Eq. (B11).

We have thus shown by examples and by general discussion that the diagonal part of a commutator $[K,H^0]$ is sometimes not zero, contrary to what routine application of the Dirac notation would suggest; the nonzeroness is caused by the nonvanishing of the bilinear concomitant; *the correct result is always obtained if we first evaluate* $KH^0 - H^0K = C$ *and then find the diagonal matrix elements of* C.

The second example given above is actually quite illustrative for the case of linear response theory, in which we met the commutator diagonal part $\mathcal{P}[A,e^{-\beta H^0}]$ for the field term. If the system were a one-electron solid subject to an electric field, then $A = e\mathbf{q}$. The commutator is found to be by direct computation, as in the example (B8),

$$[\mathbf{q},e^{-\beta H^0}] = (-i\hbar\beta/m)e^{-\beta H^0}\mathbf{p} \tag{B14}$$

[which concords with the result from Kubo's lemma (4.14)]. The matrix element involved is ($|\gamma\rangle \equiv |\mathbf{k}\rangle$):

$$\langle\gamma|\mathbf{p}|\gamma\rangle = m\langle\mathbf{v}_\mathbf{k}\rangle = (m/\hbar)\nabla_\mathbf{k}\epsilon(\mathbf{k}), \tag{B15}$$

where $\langle\mathbf{v}_\mathbf{k}\rangle$ is the expectation value of the velocity in a Bloch state and where $\epsilon(\mathbf{k})$ is the band energy of the Bloch state $|\mathbf{k}\rangle$. It is also clear that the matrix element should not vanish, for the expectation value of the current $\langle J\rangle = -e\langle\mathbf{v}_\mathbf{k}\rangle$ should be nonzero for the situation envisioned in linear response.

APPENDIX C: SOME LEMMAS

In this Appendix we prove some lemmas pertaining to the superoperators Λ_d and $i\mathscr{L}^0$, which were employed in Sec. 5–7.

Lemma 1: For any two operators C and D we have

$$\mathrm{Tr}(C\Lambda_d D) = \mathrm{Tr}(D\Lambda_d C). \tag{C1}$$

Two proofs can be given. First, we use the notion of the scalar product in Liouville space: $\{A,B\} = \mathrm{Tr}AB^\dagger$. Employing the property $(\Lambda_d D^\dagger)^\dagger = \Lambda_d D$, which is immediately verified from Eq. (2.31), we see that the statement also reads

$$\{C,\Lambda_d D^\dagger\} = \{\Lambda_d C, D^\dagger\}, \tag{C2}$$

which is satisfied since Λ_d is self-adjoint (Sec. 2.C). The second proof follows from the fact that the master operator M in function space is self-adjoint. This operator was defined by Eq. (2.13). We now have

$$\begin{aligned} C\Lambda_d D = &-\sum_{\gamma\gamma''} C|\gamma\rangle\langle\gamma|\{W_{\gamma''\gamma}\langle\gamma''|D|\gamma''\rangle \\ &- W_{\gamma\gamma''}\langle\gamma|D|\gamma\rangle\}, \end{aligned} \tag{C3}$$

from which

$$\begin{aligned} \mathrm{Tr}(C\Lambda_d D) = &-\sum_{\gamma\gamma'\gamma''}\langle\gamma'|C|\gamma\rangle\langle\gamma|\gamma'\rangle\{W_{\gamma''\gamma}\langle\gamma''|D|\gamma''\rangle \\ &- W_{\gamma\gamma''}\langle\gamma|D|\gamma\rangle\} \\ = &\int\Delta\gamma\langle\gamma|C|\gamma\rangle M\langle\gamma|D|\gamma\rangle \\ = &(\langle\gamma|C|\gamma\rangle, M\langle\gamma|D|\gamma\rangle). \end{aligned} \tag{C4}$$

Likewise, one obtains

$$\mathrm{Tr}(D\Lambda_d C) = (\langle\gamma|D|\gamma\rangle, M\langle\gamma|C|\gamma\rangle). \tag{C5}$$

Using the self-adjointness of M, as proven in Part I, Eq. (8.2), the statement follows.

Lemma 2: For any two operators C and D we have

$$\mathrm{Tr}(C\,e^{-\Lambda_d t}D) = \mathrm{Tr}(D\,e^{-\Lambda_d t}C). \tag{C6}$$

The statement follows from series expansion of the exponentials and repeated use of the lemma (C1).

Lemma 3: For any two operators C and D we have

$$\mathrm{Tr}(C\mathscr{L}^0 D) = -\mathrm{Tr}(D\mathscr{L}^0 C). \tag{C7}$$

The proof can be given directly using the commutator form for $\mathscr{L}^0 D$ and using cyclic permutativity of the trace. Or, we may notice that Eq. (6.4) is equivalent to the self-adjointness

$$\{C,\mathscr{L}^0 D^\dagger\} = \{\mathscr{L}^0 C, D^\dagger\} \tag{C8}$$

providing we note that $(\mathscr{L}^0 D^\dagger)^\dagger = -\mathscr{L}^0 D$.

Lemma 4: For any two operators C and D we have

$$\mathrm{Tr}(C\,e^{-i\mathscr{L}^0 t}D) = \mathrm{Tr}(D\,e^{i\mathscr{L}^0 t}C). \tag{C9}$$

By series expansion we have

$$\begin{aligned} &(C\,e^{-i\mathscr{L}^0 t}D) \\ &= \mathrm{Tr}\{C[1 - it\mathscr{L}^0 + \tfrac{1}{2}(it)^2\mathscr{L}^0\mathscr{L}^0 - \ldots]D\} \\ &= \mathrm{Tr}\{CD - Cit\mathscr{L}^0 D + C\tfrac{1}{2}(it)^2\mathscr{L}^0\mathscr{L}^0 D - \ldots\} \\ &= \mathrm{Tr}\{DC + Dit\mathscr{L}^0 C + D\tfrac{1}{2}(it)^2\mathscr{L}^0\mathscr{L}^0 C + \ldots\} \\ &= \mathrm{Tr}\{D[1 + it\mathscr{L}^0 + \tfrac{1}{2}(it)^2\mathscr{L}^0\mathscr{L}^0 + \ldots]C\} \\ &= \mathrm{Tr}(D\,e^{i\mathscr{L}^0 t}C); \end{aligned} \tag{C10}$$

the changing of the terms involves repeated application of Lemma 2 and application of the cyclic permutativity property of the trace, e.g., for the cubic term

$$\begin{aligned} \mathrm{Tr}(C\mathscr{L}^0\mathscr{L}^0\mathscr{L}^0 D) &= -\mathrm{Tr}[(\mathscr{L}^0\mathscr{L}^0 D)\mathscr{L}^0 C] \\ &= -\mathrm{Tr}[(\mathscr{L}^0 C)\mathscr{L}^0\mathscr{L}^0 D] \\ &= +\mathrm{Tr}[(\mathscr{L}^0 D)\mathscr{L}^0\mathscr{L}^0 C] \\ &= +\mathrm{Tr}[(\mathscr{L}^0\mathscr{L}^0 C)\mathscr{L}^0 D] \\ &= -\mathrm{Tr}(D\mathscr{L}^0\mathscr{L}^0\mathscr{L}^0 C). \end{aligned}$$

Lemma 5: We have

$$\mathrm{Tr}(C\,e^{(\Lambda_d + i\mathscr{L}^0)t}D) = \mathrm{Tr}(D\,e^{(\Lambda_d - i\mathscr{L}^0)t}C). \tag{C11}$$

This lemma follows from the preceding ones, for we have

$$\begin{aligned} \mathrm{Tr}[C\,e^{(\Lambda_d + i\mathscr{L}^0)t}D] = \mathrm{Tr}\{C[1 &+ (\Lambda_d + i\mathscr{L}^0)t + \tfrac{1}{2}t^2 \\ &\times(\Lambda_d + i\mathscr{L}^0)(\Lambda_d + i\mathscr{L}^0) + \ldots]D\} \end{aligned}$$

$$= \mathrm{Tr}\{C\,[1 + (\Lambda_d + i\mathscr{L}^0)t + \tfrac{1}{2}t^2(\Lambda_d\Lambda_d + \Lambda_d i\mathscr{L}^0 + i\mathscr{L}^0\Lambda_d - \mathscr{L}^0\mathscr{L}^0) + \ldots]D\}$$

$$\overset{\text{lemmas 1,3}}{=} \mathrm{Tr}\{D\,[1 + (\Lambda_d - i\mathscr{L}^0)t + \tfrac{1}{2}t^2(\Lambda_d\Lambda_d - i\mathscr{L}^0\Lambda_d - \Lambda_d i\mathscr{L}^0 - \mathscr{L}^0\mathscr{L}^0) + \ldots]C\}$$

$$= \mathrm{Tr}\{D\,[1 + (\Lambda_d - i\mathscr{L}^0)t + \tfrac{1}{2}t^2(\Lambda_d - i\mathscr{L}^0)(\Lambda_d - i\mathscr{L}^0) + \ldots]C\} = \mathrm{Tr}[D\,e^{(\Lambda_d - i\mathscr{L}^0)t}C]. \quad \text{(C12)}$$

APPENDIX D: ON FERMION OPERATORS

The usual results shown in the textbooks are

$$c_\xi^\dagger|n_1 n_2 \ldots 0_\xi \ldots\rangle = (-1)^{\sum(1,\xi-1)}|n_1 n_2 \ldots 1_\xi \ldots\rangle, \quad \text{(D1)}$$

$$c_\xi|n_1 n_2 \ldots 1_\xi \ldots\rangle = (-1)^{\sum(1,\xi-1)}|n_1 n_2 \ldots 0_\xi \ldots\rangle, \quad \text{(D2)}$$

with $c_\xi^\dagger|\ldots 1_\xi \ldots\rangle = 0$ and $c_\xi|\ldots 0_\xi \ldots\rangle = 0$. The fuller results of Eqs. (8.11) and (8.12), necessary for the present paper, are easily derived from the basic commutation rules

$$c_\zeta\, c_\xi^\dagger + c_\xi^\dagger\, c_\zeta = \delta_{\zeta\xi}, \quad \text{(D3)}$$

$$c_\zeta c_\xi + c_\xi c_\zeta = 0, \quad \text{(D4)}$$

$$c_\xi^\dagger c_\xi^\dagger + c_\xi^\dagger c_\xi^\dagger = 0. \quad \text{(D5)}$$

From Eq. (D3) we have

$$(c_\zeta c_\xi^\dagger)(c_\zeta c_\xi^\dagger) = c_\zeta\,(c_\xi^\dagger c_\zeta)c_\xi^\dagger = c_\zeta\,(1 - c_\zeta c_\xi^\dagger)c_\xi^\dagger = c_\zeta c_\xi^\dagger \quad \text{(D6)}$$

since by Eq. (D5) $c_\xi^\dagger c_\xi^\dagger = 0$. The operator $c_\zeta c_\xi^\dagger$ therefore has eigenvalues 0 and 1, so by convention we identify it with the operator $1 - \mathbf{n}_\zeta$; hence, $\mathbf{n}_\zeta = c_\xi^\dagger c_\zeta$. Denoting the eigenvalues of the operator $\mathbf{n}_\zeta$ by n_ζ, we have denoting by $|\{n\}\rangle$ an occupation number state

$$\mathbf{n}_\zeta|\{n\}\rangle = n_\zeta|\{n\}\rangle, \quad \text{(D7)}$$

or dotting into $\langle\{n\}|$:

$$\langle\{n\}|c_\xi^\dagger\, c_\zeta|\{n\}\rangle = n_\zeta, \quad \text{(D8)}$$

showing that $c_\zeta|\{n\}\rangle$ is normalized to n_ζ. Similarly, we find

$$\langle\{n\}|c_\zeta\, c_\xi^\dagger|\{n\}\rangle = 1 - n_\zeta, \quad \text{(D9)}$$

showing $c_\xi^\dagger|\{n\}\rangle$ is normalized to $1 - n_\zeta$. Considering now the repeated operations $\mathbf{n}_\zeta\, c_\xi^\dagger$ and $(1 - \mathbf{n}_\zeta)c_\zeta$, we have

$$\mathbf{n}_\zeta\, c_\xi^\dagger|\{n\}\rangle = c_\xi^\dagger(c_\zeta\, c_\xi^\dagger)|\{n\}\rangle = c_\xi^\dagger(1 - \mathbf{n}_\zeta)|\{n\}\rangle = (1 - n_\zeta)c_\xi^\dagger|\{n\}\rangle, \quad \text{(D10)}$$

and

$$(1 - \mathbf{n}_\zeta)c_\zeta|\{n\}\rangle = c_\zeta\,(c_\xi^\dagger\, c_\zeta)|\{n\}\rangle = n_\zeta\, c_\zeta\{n\}\rangle. \quad \text{(D11)}$$

Subtracting from both sides $c_\zeta|\{n\}\rangle$, this gives also

$$\mathbf{n}_\zeta\, c_\zeta|\{n\}\rangle = (1 - n_\zeta)c_\zeta|\{n\}\rangle. \quad \text{(D11a)}$$

According to Eq. (D10), the state $c_\xi^\dagger|\{n\}\rangle$ is an eigenstate of $\mathbf{n}_\zeta$ with eigenvalue $1 - n_\zeta$. Hence, apart from a phase factor, $c_\xi^\dagger|\{n\}\rangle = \text{const.}|\ldots 1 - n_\zeta \ldots\rangle$. The constant follows from the normalization statement derived above. Thus

$$c_\xi^\dagger|\{n\}\rangle = (-1)^\alpha\sqrt{1 - n_\zeta}\,|n_1 n_2 \ldots 1 - n_\zeta \ldots\rangle. \quad \text{(D12)}$$

Likewise, from Eq. (D11a) it follows that $c_\zeta|\{n\}\rangle = \text{const.}|\ldots 1 - n_\zeta \ldots\rangle$. With the normalization indicated by Eq. (D8) we find

$$c_\zeta|\{n\}\rangle = (-1)^\beta\sqrt{n_\zeta}\,|n_1 n_2 \ldots 1 - n_\zeta \ldots\rangle. \quad \text{(D13)}$$

The parameters α and β are fixed by requiring Eqs. (D4) and (D5) for $\zeta \neq \xi$, and by considering the repeated operations $c_\zeta c_\xi|\{n\}\rangle$ and $c_\xi^\dagger c_\xi^\dagger|\{n\}\rangle$. This leads to $\alpha = \beta = \Sigma(1,\zeta - 1)$, where $\Sigma(1,\zeta)$ dentoes the occupancies of all ordered one-particle states $|1)\ldots|\zeta)$. Thus, we obtain the precise forms of Eqs. (8.11) and (8.12). The results are now in a form entirely analogous to those for boson operators.

It may be argued that for fermions the square root signs make no sense since n_ζ and n_ζ^2 have the same store of values (0 and 1). However, only by maintaining the square root signs are we led to the Boltzmann equation in its simplest form which now is, moreover, very similar for fermions and bosons.

There is another problem associated with fermion operators in a canonical ensemble. The following rule was shown by Fowler[24] using the Darwin–Fowler method:

$$\langle\Delta n_\zeta \Delta n_\xi\rangle_{\text{eq}} = -\beta\frac{\partial\langle n_\zeta\rangle_{\text{eq}}}{\partial e_\xi} = \langle n_\zeta\rangle_{\text{eq}}(1 - \langle n_\zeta\rangle_{\text{eq}})\left(\delta_{\zeta\xi} - \frac{\partial\mu}{\partial e_\xi}\right), \quad \text{(D14)}$$

where e_ξ is the energy of a state $|\xi)$ and μ is the electrochemical potential or Fermi level. (A similar relationship holds for bosons, replacing $1 - \langle n_\zeta\rangle_{\text{eq}}$ by $1 + \langle n_\zeta\rangle_{\text{eq}}$.) For $\zeta \neq \xi$, the relation gives no problems; it then only indicates that the cross correlation is not zero as in the grand canonical ensemble. However, for $\zeta = \xi$, it follows that $\langle\Delta n_\zeta^2\rangle_{\text{eq}} \neq \langle n_\zeta\rangle_{\text{eq}} - \langle n_\zeta\rangle_{\text{eq}}^2$, which entails that $\langle n_\zeta^2\rangle_{\text{eq}} \neq \langle n_\zeta\rangle_{\text{eq}}$. This is certainly an anomaly. Whereas the correction term is very small for the microscopic variance ($\partial\mu/\partial e_\xi$ is of order $1/\mathscr{N}$, where $\Sigma n_\zeta = \mathscr{N}$ is the canonical constraint), the correction becomes large for coarse grained occupancies $N_k = \Sigma_{\zeta=1}^{Z_k} n_\zeta$, when Z_k is of order $\mathscr{N}$. Equation (D14) with correction factor predicts correctly the noise $\langle\Delta N_k^2\rangle_{\text{eq}}$ of coarse grained occupancies such as the fluctuation of all electrons in the conduction band of a semiconductor.

[1] K.M. van Vliet, J. Math. Phys. **19**, 1345 (1978).
[2] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[3] L. van Hove, Physica (Utrecht) **21**, 517 (1955).
[4] D_γ can also act on any nondiagonal operator; it then destroys the nondiagonal part: $D_\gamma K = \langle\gamma|K|\gamma\rangle = \langle\gamma|K_d|\gamma\rangle$.
[5] L. van Hove, in *Les Gaz Neutres et Ionises*, edited by C. de Witt and J.F. Detoeuf (Hermann, Paris, 1960), p. 159.
[6] Occasionally other boundary conditions are encountered, as for the Hermite polynomials or associated Laguerre polynomials for magnetic phenomena. Since these involve an infinite domain, they usually present no boundary term problems.
[7] U. Fano, *Lectures on the Many-Body Problem*, edited by E.R. Caianiello (Academic, New York, 1964), Vol. 2.
[8] As in Paper I, a bracketed operator $A(t)$ is a time dependent operator, either of the Heisenberg type $A^H(t)$, the reduced type $A^R(t)$, or of the interaction type $A^I(t)$. Unbracketed operators are time-independent, and, if non-superscripted, they are Schrödinger operators A^S. Naturally $A^H(0) = A^I(0) = A^S = A$.
[9] I owe this remark to M. Dresden, Vosbergen Conference, 1968.

[10]In a previous version of this manuscript we solved for the Green's function of (3.4), defined by

$$\frac{\partial \mathcal{K}(t,t')}{\partial t} + i(1-\mathcal{P})\mathcal{L}\mathcal{K}(t,t') - \frac{i}{\hbar}F(t)(1-\mathcal{P})[A,\mathcal{K}(t,t')] = \mathcal{I}\delta(t-t').$$

The solution is

$$\mathcal{K}(t,t') = u(t-t')e^{i(t-t')(1-\mathcal{P})\mathcal{L}} \times \exp\left[i\int_{t'}^{t} d\tau F(\tau)(1-\mathcal{P})\mathcal{C}_A(t'-\tau)\right]$$

where $\mathcal{C}_A$ is a superoperator such that

$$\mathcal{C}_A(t)B = (1/\hbar)e^{it'(1-\mathcal{P})\mathcal{L}}[A,e^{it(1-\mathcal{P})\mathcal{L}}B].$$

While this solution separates more completely the diagonal and nondiagonal parts, it is of little value, for the linearized equations reduce again to (3.11) and (3.12).

[11]R. Zwanzig, in *Lectures in Theoretical Physics, Vol, III*, edited by W.E. Britten and J. Downs, Boulder, Colorado, 1960 (Interscience, New York, 1961), pp. 106–141.

[12]ρ is bounded; for arbitrary ϕ of $\mathcal{H}$ we have

$$(\rho\phi,\rho\phi) = \sum_{\gamma} |(\phi,\gamma)|^2(\rho^2\gamma,\gamma) < \sum_{\gamma} |(\phi,\gamma)|^2(\rho\gamma,\gamma) \leqslant \mathrm{Tr}\,\rho = 1.$$

[13]We prefer the notation J_A over $\dot{A}^R$. However, to keep the equations in Kubo-form appearance, we mainly use the notation $\dot{A}^R$ in Secs. 6 and 7.

[14]In Paper I we computed $B^R_{ll}(\Delta t)$, where we assumed from the outset that B was diagonal. If this assumption is dropped, then I, Eq. (6.11) will read $\langle\gamma|U^{(0)\dagger}BU^{(0)}|\gamma\rangle = \langle\gamma|B^I(\Delta t)|\gamma\rangle$. In the higher order terms the diagonal approximation is allowed. With this modification, I, (6.37) will read $B^R(\Delta t) = B^I(\Delta t) - \Lambda_{ll}B_{ll}\Delta t$. By subtracting from both sides $B(0)$, dividing by Δt, and operating with $\mathcal{P}$, the statement (4.28) follows if we set $B \to A$. In principle, we can also obtain (4.28) from the van Hove limit of the Heisenberg equation of motion.

[15]We assume that B is a vector of $\mathcal{H}$, i.e., translationally invariant over the dimensions of the system.

[16]H. Grad, in *Encyclopedia of Physics* (Springer, New York, 1958) Vol. 12.

[17]L. Boltzmann *Lectures on Gas Theory* (University of California, Berkeley, 1964)[translated from *Vorlesungen uber Gastheorie* (Barth, Leipzig, 1896, 1898)].

[18]See, for example, A.H. Kahn and H.P.R. Frederickse, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1959), Vol. 9.

[19]D. Kennedy, C.-T. Hsing, A. Sutherland, and K.M. van Vliet, Phys. Status Solidi (a), in press; also, Chi-Tien Hsing, Ph.D. Thesis, University of Florida (1977).

[20]M. Charbonneau and K.M.van Vliet, to be published.

[21]See, for example, C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).

[22]K.M. van Vliet, Phys. Status Solidi (b) **78**, 667 (1976).

[23]F. Riesz and B. Sz. Nagy, *Leçons d'analyse Fonctionelle* (Gauthiers-Villars, Paris, 1965), 4th edition.

[24]R.H. Fowler, *Statistical Mechanics* (Cambridge University, Cambridge, England, 1936) [(reprinted by Dover, New York, 1966)]; Eqs. (2114) and (2115).

All spontaneously broken symmetries for noncovariant currents

W. D. Garber[a), b)]

Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08540
(Received 5 December 1978)

For a current not covariant under translations, the associated symmetry can be spontaneously broken even in the case of a mass gap. The most general form of such a symmetry in a Wightman field theory is determined; it is shown that the symmetry generator can be evaluated on scattering states, commutes (as a form) with the S matrix, and has vanishing matrix elements between states whose particle number differs by more than one.

1. INTRODUCTION

In this paper, the study of symmetries induced by local, conserved currents (in the framework of a Wightman quantum field theory) is continued. In order to clarify the various assumptions made in the literature about such currents, it seems helpful to consider the following four examples:

$$(1)\quad j^{\mu}{}_{\nu}(x) := :\phi_m \partial^{\mu}\partial_{\nu}\phi_m - (\partial^{\mu}\phi_m)\partial_{\nu}\phi_m:(x), \tag{1.1}$$

$$(2)\quad j^{\mu}{}_{\nu\kappa}(x) := x_{\nu} j^{\mu}{}_{\kappa}(x) - x_{\kappa} j^{\mu}{}_{\nu}(x), \tag{1.2}$$

$$(3)\quad j^{\mu}(x) := c\partial^{\mu}\phi_0(x), \tag{1.3}$$

$$(4)\quad j^{\mu}(x) := h(x)\partial^{\mu}\phi_m(x) - (\partial^{\mu}h)(x)\phi_m(x). \tag{1.4}$$

Here $\phi_m(x)$ denotes the free scalar field of mass $m \geqslant 0$, the double dots denote Wick ordering, the function $h(x)$ in example (4) (which is due to Ref. 1), is asumed to be a real solution [in $O_M(\mathbb{R}^4)$] of the Klein–Gordon equation for the same mass m as the field $\phi_m(x)$, and c in example (3) is a real constant; in example (2), $j^{\mu}{}_{\nu}$ denotes the current from example (1).

All four currents are Hermitian, conserved, local, and relatively local with respect to the field ϕ_m from which they are constructed. The associated generator, which one expects to be

$$Q = \int j^0(x) d^3x$$

(all indices other than μ having been suppressed), is ill-defined by this equation. However, considering "approximate generators"

$$Q_r := j^0(\vartheta_r \eta), \tag{1.5}$$

with

$$\vartheta_r(\mathbf{x}) := \vartheta(r^{-1}|\mathbf{x}|), \quad \vartheta \in \Theta,$$

$$\Theta := \left\{ \vartheta \in \mathscr{D}([0,\infty])/\vartheta(s) : \begin{cases} 1, & 0 \leqslant s \leqslant 1 \\ 0, & s \geqslant 2 \end{cases} \right\}, \tag{1.6}$$

$$\eta \in H, \quad H := \left\{ \eta \in \mathscr{D}(\mathbb{R}^1) \Big/ \int \eta(x^0) dx^0 = 1 \right\} \tag{1.7}$$

and defining the generator Q on states A [A: any polynomial in operators $\phi_m(f)$ with $f \in \mathscr{D}(\mathbb{R}^4)$; Ω: vacuum state] by

$$QA\Omega := \lim_{r \to \infty} [Q_r, A]\Omega, \tag{1.8}$$

we readily see that the action of Q in the four examples is

$$(1)'\quad i[Q_{\nu}, \phi_m(x)] = \partial_{\nu}\phi_m(x), \tag{1.9}$$

$$(2)'\quad i[Q_{\nu\kappa}, \phi_m(x)] = (x_{\nu}\partial_{\kappa} - x_{\kappa}\partial_{\nu})\phi_m(x), \tag{1.10}$$

$$(3)'\quad i[Q, \phi_0(x)] = c, \tag{1.11}$$

$$(4)'\quad i[Q, \phi_m(x)] = h(x). \tag{1.12}$$

The first two examples are thus the familiar generators of the Poincaré group, and in the last two examples, a constant c resp. a function $h(x)$ (which is a solution of the Klein–Gordon equation) is added to the field.

The currents in examples (1) and (3) have the additional property that they are translationally covariant:

$$U(a) j^{\mu}(x) U(-a) = j^{\mu}(x + a) \tag{1.13}$$

[where $U(a)$, $a \in \mathbb{R}^4$, is the unitary representation of the translation group]. Under the additional assumption of a mass gap, the structure of the generator Q corresponding to an arbitrary translationally covariant current constructed from interacting fields has been obtained in Ref. 2: Q can be shown to be a symmetric (even self-adjoint) operator defined on the asymptotic (free) fields and acting on them as a polynomial in the linear momentum operator. Thus, a slight generalization of (1.9) (replacing ∂_{ν} by a polynomial in ∂_{κ}, $\kappa = 0,1,2,3$) is all that can be obtained from such currents. The fact that Q is symmetric is equivalent to

$$\lim_{r \to \infty} (\Omega \mid [Q_r, A]\Omega) = 0, \tag{1.14}$$

where Q_r, A, Ω are as in (1.8), i.e., to the fact that the symmetry is not spontaneously broken. For a translationally covariant current, this is implied by a mass gap being present. Thus, in order to investigate spontaneously broken symmetries in this case, one has to drop the mass gap assumption. Hence, it is no accident that in example (3) for which

$$\lim_{r \to \infty} (\Omega \mid [Q_r, \phi_0(x)]\Omega) = -ic$$

a mass zero field appears.

Examples (2) and (4) show that the assumption of translational covariance is, in fact, too restrictive, ruling out even

[a)] On leave of absence from the Institute für Theoretische Physik, Universatät Göttingen, Bunsenstr. 9, Göttingen, Germany.
[b)] Supported by a DAAD grant.

 0022-2488/79/122596-09$01.00

the generators of the Poincaré group. It seems therefore natural to investigate arbitrary translationally noncovariant currents. However, now one can have a spontaneously broken symmetry in a theory with mass gap, as example (4) with $m > 0$ shows, for which

$$\lim_{r\to\infty}(\Omega\,|\,[Q_r,\phi_m(x)]\Omega) = -ih(x). \tag{1.15}$$

The structure of unbroken symmetries for noncovariant currents has been investigated in Refs. 3 and 4 under the assumption of a mass gap: The associated generator Q is again symmetric [by (1.14)], defined on the asymptotic fields, and acts on them as a polynomial in x_κ and ∂_ν, κ, $\nu = 0,1,2,3$. Thus, a slight generalization of examples (1) and (2) is all that can be obtained from an arbitrary noncovariant current, under the above assumptions.

Since example (4) shows that there are cases of spontaneously broken symmetries in theories with a mass gap, one might inquire about the structure of such symmetries. It seems easier to ask that question in theories with a mass gap for which there are no infrared problems and the Haag–Ruelle scattering theory is available. However, it must be emphasized that no result for the translationally covariant case can be obtained by simply specializing the results of the present paper, as in the case of unbroken symmetries. In the following, the existence of a mass gap is assumed, but (1.14) need not hold.

There is an immediate difficulty if (1.14) is dropped: An operator Q could still be defined by (1.8), but since (1.14) does not hold, it is by necessity not a symmetric operator and as such has no chance of being the generator of a symmetry. Hence, the associated generator must be defined by another method. Since one is primarily interested not in the operator, but in its expectation values, a definition as a quadratic form seems even more natural:

$$Q(\varphi\,|\,\psi) := \lim_{r\to\infty}(\varphi\,|\,Q_r\psi). \tag{1.16}$$

Note that (1.8) and (1.16) do not agree for the examples (3), (4) of spontaneously broken symmetries. As no commutator is employed in the definition (1.16), one seems to be unable to use locality effectively; in particular, it is not even clear if Q is defined on the states $A\Omega$ used in (1.8). Nevertheless, (1.16) will be adopted as the definition of Q in the following, and it will be shown (in Sec. 3) that Q can be defined as a quadratic form even on (the usual dense set, cf. Refs. 3, 4, of) asymptotic scattering states, commutes (as a form) with the S matrix and has a simple form on asymptotic states: It consists of a term bilinear in annihilation and creation operators, familiar from Ref. 3, to which is added a term linear in these operators: the last term is responsible for the symmetry breaking. In particular, Q vanishes on asymptotic states ϕ, ψ for which the particle number differs by more than one. Section 2 contains some preliminary estimates.

2. SOME PROPERTIES OF Q

In this section, the assumptions on the current will be described in detail, and simple properties of Q will be derived.

Consider a Wightman quantum field theory given by a finite set of fields $\{\phi_i(x)\}$ with common dense domain D (stable and Lorentz invariant), vacuum vector Ω, and translations $U(a)$. In addition to the translationally covariant fields $\{\phi_i(x)\}$, there is assumed to be given a noncovariant current density $J^\mu(x)$ with the properties

(J1) $J^\mu(x)$ is an operator-valued tempered distribution, defined on D and leaving D invariant;

(J2) $J^\mu(x)$ is Hermitian;

(J3) $J^\mu(x)$ is local and relatively local with respect to the fields $\{\phi_i(x)\}$;

(J4) $J^\mu(x)$ is conserved:

$$\partial_\mu J^\mu(x) := \frac{\partial}{\partial x^\mu}J^\mu(x) = 0.$$

[In (J4), summation over the repeated index μ is implied.] The tranlates of J^μ,

$$J^\mu(x,a) := U(a)\,J^\mu(x)U(-a) \tag{2.1}$$

are, by a regular transformation, distribution of $x + a$ and x,

$$j^\mu(x + a, x) := J^\mu(x,a) \tag{2.2}$$

and have the following properties, which follow immediately from (J1)–(J4) and (2.1), (2.2):

(j0) $U(a)j^\mu(y, x)U(-a) = j^\mu(y + a, x)$;

(j1) $j^\mu(y,x)$ is an operator valued tempered distribution in y and x, defined on D and leaving D invariant;

(j2) $j^\mu(y,x)$ is Hermitian;

(j3) $j^\mu(y,x)$ is local and relatively local with respect to the fields $\{\phi_i\}$ in the variable y, for all x;

(j4) $j^\mu(y,x)$ is conserved:

$$\left(\frac{\partial}{\partial y^\mu} + \frac{\partial}{\partial x^\mu}\right)j^\mu(y,x) = 0\,. \tag{2.3}$$

So far, the assumptions are the same as those made in Ref. 4. As in Ref. 4, the following normalization condition is assumed:

(j5) $(\Omega\,|\,j^\mu(y,x)\Omega) = 0$,

[equivalent to

(J5) $(\Omega\,|J^\mu(x)\Omega) = 0$],

which can always be achieved by redefining

$$\hat{j}^\mu(y,x) := j^\mu(y,x) - (\Omega\,|\,j^\mu(y,x)\Omega); \tag{2.4}$$

if j^μ fulfills (j0)–(j4), $\hat{j}^\mu$ fulfills (j0)–(j5). However, unlike the case of unbroken symmetries, where Q, as defined in (1.8), is not affected by this normalization, the quadratic form Q defined in (1.16) changes under (2.4). Nevertheless, (j5) will be assumed as it has the physically reasonable consequence

$$Q(\Omega\,|\,\Omega) = 0. \tag{2.5}$$

Denote by P_{SL} the algebra of strictly local operators generated by $\phi_i(x)$ smeared with test functions from $\mathscr{D}(\mathbb{R}^4)$. It will next be shown that the limit (1.14) exists even though it need not be zero:

2.1 *Lemma*: For fixed $\psi,\phi\in D$, the functional

$$q_r(A) := (\psi|\,[Q_r\,,A\,]\phi) \tag{2.6}$$

has a limit, as $r\to\infty$, for all $A\in P_{\mathrm{SL}}$. This limit, denoted by

$q(A)$, is independent of $\vartheta\in\Theta$ and $\eta\in H$, as defined in (1.6) and (1.7).

Whenever the dependence of q_r on the vectors ψ,ϕ, is important, this functional will be denoted by $q_r(A;\psi,\phi)$. Otherwise these vectors are supposed to be chosen and kept fixed.

Proof: By locality, the limit is reached for finite $r = r_0(A)$ and thus independent of $\vartheta\in\Theta$ [cf. (1.6)]. Choose $\eta_1,\eta_2\in H$ [cf. (1.7)] and consider

$$\eta(x^0) = \int_{-\infty}^{x^0} [\eta_1(y^0) - \eta_2(y^0)]dy^0,$$

which has compact support by the definition of H. By current conservation, for $r > r_0(A)$,

$$(\psi|[j^0(\vartheta_r\cdot(\eta_1 - \eta_2)),A]\phi) = (\psi|[j^0(\vartheta_r\cdot\partial_0\eta),A]\phi)$$
$$= (\psi|[j^i(\partial_i\vartheta_r\cdot\eta),A]\phi) = 0,$$

since ϑ_r is constant on the support of A. □

By the above lemma, for the translates $A(x) := U(x)AU(-x)$,

$$\lim_{r\to\infty} q_r(A(x)) = q(A(x))$$

for each fixed x and $A\in P_{\mathrm{SL}}$. It follows from the assumptions that $q_r(A(x))$ is, in fact, an O_M function in x for fixed A. As a function in O_M, it defines a tempered distribution. It will next be shown that $q(A(x))$ is a tempered distribution also and that $q_r\to q$ in the topology of $\mathscr{S}'$. A similar result was obtained in Ref. 5 for a restricted class of currents; the same proof holds, however, in general, and a slight change even gives the rate of convergence:

2.2 *Lemma*: For all $A\in P_{\mathrm{SL}}$ and all positive integers N, there is some $\mathscr{S}(\mathbb{R}^4)$-norm $\|\cdot\|_{m,n}$ with $m = N + K$ and K,n constant such that, for all $f\in\mathscr{S}(\mathbb{R}^4)$,

$$|q_r(A(f)) - q(A(f))| \leqslant C_N(1+r)^{-N}\|f\|_{N+K,n}, \tag{2.7}$$

where

$$A(f) = \int A(x)f(x)\,dx$$

and the constant C_N depends on N,A,K and n, but not on r.

The precise definition of $\|\cdot\|_{m,n}$ is given in the proof below [cf. (2.13)].

Proof: Choose fixed vectors $\phi,\psi\in D$, cf. (2.6). By (J1),

$$C(x,y) := (\phi|[J^0(x),A(y)]\psi) \tag{2.8}$$

is a tempered distribution in x and y; hence, there are integers s,t with

$$|C(g)| \leqslant C_{s,t}\|g\|_{s,t} \tag{2.9}$$

for all $g = g(x,y)\in\mathscr{S}(\mathbb{R}^8)$. By (J3), $C(x,y) = 0$ whenever $x - y$ is spacelike to the support of A, i.e., in particular if $x - y\in H_b := \{z\in\mathbb{R}^4 \mid |z_0| - \|\mathbf{z}\| < -b\}$ for b large enough, where the size of b depends only on (the support of) A. Keep that b fixed. Let θ be a function in O_M which is zero on H_b and 1 on the complement of H_{b-1} (choose $b > 1$). Then

$$C(x,y) = C(x,y)\theta(x-y). \tag{2.10}$$

By (2.9) and the definition of q_r and q, it is thus enough to prove, for some m,n depending on s,t, and N,

$$\|[\vartheta_r(\mathbf{x}) - 1]\eta(x^0)\theta(x-y)f(y)\|_{s,t} \leqslant C_N(1+r)^{-N}\|f\|_{m,n}. \tag{2.11}$$

This will now be proven for *all* s,t which implies, in particular, that $y(x^0)\theta(x-y)f(y)$ is in $\mathscr{S}(\mathbb{R}^8)$ so that $q(A(f))$ is defined. Denote the function on the left side of (2.11) by $h_r(x,y)$.. Then $h_r\neq 0$ at most for $\|\mathbf{x}\|\geqslant r$ and $x - y$ in the complement of H_b, i.e., for $|x_0 - y_0| - \|\mathbf{x} - \mathbf{y}\| > -b$ which implies

$$|y| := |y_0| + \|\mathbf{y}\| \geqslant |y_0 - x_0| - |x_0| - \|\mathbf{y} - \mathbf{x}\| + \|\mathbf{x}\|$$
$$\geqslant -|x_0| - b + \|\mathbf{x}\| \geqslant -d - b + \|\mathbf{x}\| \tag{2.12}$$
$$\geqslant \|\mathbf{x}\| - r/2 \geqslant \tfrac{1}{2}\|\mathbf{x}\|,$$

where the support of η lies in $[-d,d]$, and r has been chosen larger than $2(d+b)$. Now in

$$\|h_r(x,y)\|_{s,t} := \sup_{\substack{|\alpha|\leqslant s\\ |\beta|\leqslant t}} \Big| |y|^{\alpha}|x|^{\alpha}D_y^{\beta}D_x^{\beta}[(\vartheta_r(\mathbf{x}) - 1)\eta(x_0)f(y) \times\theta(x-y)]\Big| \tag{2.13}$$

the $\mathbf{x}$ derivatives are bounded by a polynomial in $\mathbf{x}$ and $\mathbf{y}$ independent of r; the x_0 derivatives are bounded by a polynomial in x_0 and y_0. Hence $[\alpha = (\alpha_0,\boldsymbol{\alpha})]$:

$$\|h_r(x,y)\|_{s,t} \leqslant C\sup|\,|y|^{\alpha+\gamma}|x|^{\alpha+\gamma}[D_y^{\beta}f(y)]\theta(x-y)|$$
$$\leqslant C'\|f\|_{s+|\alpha|+|\gamma|+|\gamma|,t}$$

by (2.11) and $|x_0| < d$. Lastly

$$\|h_r(x,y)\|_{s,t} \leqslant \Big(\sup_{(x,y)\in\operatorname{supp}h_r} |x|^{-N}|y|^{-N}\Big)\|h_r\|_{N+s,t}$$
$$\leqslant C_N\gamma^{-N}\|h_r\|_{N+s,t}$$

since $2|y| > \|\mathbf{x}\| > r$, by (2.12), which establishes (2.11) with $m = N + |\gamma| + s + |\alpha| + |\gamma|$, $n = t$. Now choose the particular values of (2.9) for s and t which implies (2.7) with $K = s + \alpha + |\gamma| + |\gamma|$. □

In the construction of Haag–Ruelle scattering states, the algebra P_{QL} of operators of the form

$$A(f) := \int A_1(x_1)\cdots A_l(x_l)f(x_1,\ldots,x_l)d^4x_1\cdots d^4x_l,$$
$$\forall l, f\in\mathscr{S}(\mathbb{R}^{4l}),\ A_i\in P_{\mathrm{SL}}, \tag{2.14}$$

is used. Using the technique of the preceding lemma, it can also be shown that $q_r(B)$ converges even for $B\in P_{\mathrm{QL}}$; the limit will again be denoted by $q(B)$:

Corollary 2.3: Let $A(f)$ be defined by (2.14), and $\psi,\phi\in D$. Then for all positive integers N, there is some $\mathscr{S}(\mathbb{R}^{4l})$-norm $\|\ \|_{N+K,n}$ with K,n constant such that, for all $f\in\mathscr{S}(\mathbb{R}^{4l})$,

$$|q_r(A(f);\psi,\phi) - q(A(f);\psi,\phi)| \leqslant C_N^{\psi,\phi}(1+r)^{-N}\|f\|_{N+K,n}. \tag{2.15}$$

If, in particular, $\psi = B_1(g_1)\Omega, \phi = B_2(g_2)\Omega$ with $B_1(g_1)$, $B_2(g_2)\in P_{\mathrm{QL}}$, then

$$C_N^{\psi,\phi} \leqslant C_N\|g_1\|_{a,b}\|g_2\|_{a',b'}$$

where the $\mathscr{S}$-norms $\|\ \|_{a,b}$, $\|\ \|_{a',b'}$ are independent of N.

Proof: Since

$$q_r(A(f);\psi,\varphi) = \int \sum_{i=1}^{l} q_r(A(x_i);\psi_i,\varphi_i)f(x_1,\ldots,x_l)d^4x_1\cdots d^4x_l$$

with

$$\psi_i := A^*_{i-1}(x_{i-1})\cdot\cdots\cdot A^*_1(x_1)\psi,$$
$$\varphi_i := A_{i+1}(x_{i+1})\cdot\cdots\cdot A_l(x_l)\varphi,$$

one can reduce the proof to that of the last lemma by considering

$$C_i(x_1,\ldots,x_l;x) = (\psi_i\,|\,[J^0(x),A_i(x_i)]\varphi_i)$$

which is a distribution in all variables x; $x_1,\ldots,x_l$ and hence

$$|C_i(g)| \leqslant K_i \|g\|_{s(i),t(i)} \tag{2.16}$$

for some $\mathscr{S}-(\mathbb{R}^{4(l+1)}$-norm $\|\cdot\|_{s(i),t(i)}$. Now replace, as before, g by

$$h_r(x_1,\ldots,x_l;x) := [\vartheta_r(\mathbf{x})-1]\eta(x^0)\theta(x-x_i)f(x_1,\ldots,x_l),$$

and sum (2.16) over i to get

$$|q_r(A(f))-q(A(f))| \leqslant C\|h_r\|_{s,t},$$

where $s := \max s(i)$, $t := \max t(i)$, and $C := \max C_i$. Now use (2.12) to bound $\|h_r\|_{s,t}$ as stated in (2.15).

If $\psi = B_1(g_1)\Omega$, $\varphi = B_2(g_2)\Omega$, simply note that C_i is now a distribution also in its dependence on g_1, g_2; thus (2.16) can be replaced by

$$|C_i(g;g_1,g_2)| \leqslant \hat{K}_i\|g\|_{s(i),t(i)}\cdot\|g_1\|_{a,b}\|g_2\|_{a',b'}. \square$$

The last corollary turns out to be useful in proving that the quadratic form Q of (1.16) is defined on $P_{\mathrm{QL}}\Omega \times P_{\mathrm{QL}}\Omega$. The crucial fact linking Q with q_r is part (ii) of the following lemma which is a generalization of Lemma 4.1 of Ref. 2 due to Reeh; part (i) is Lemma III of Haag, Kastler, and Robinson[6] which critically uses the mass gap assumption.

2.4 *Lemma*: (i) For any $B\in P_{\mathrm{QL}}$, there exists $\tau(B)\in P_{\mathrm{QL}}$ with

$$B\Omega - (\Omega\,|B\Omega)\Omega = P_0\tau(B)\Omega \tag{2.17}$$

(P_0: generator of time translations); furthermore, τ is a linear functional on P_{QL} preserving hermiticity.

(ii) For any $B\in P_{\mathrm{QL}}$,

$$2(\Omega\,|BQ_r\,\Omega) = (\Omega\,|\,[Q_r,B\,]\Omega) + (\Omega\,|\,[[Q_r,P_0\,],\tau(B)]\Omega) \tag{2.18}$$

Proof: (i) Consider the spectral decomposition

$$U(x) = \int e^{ipx}\,dE(p) + E_0\,, \tag{2.19}$$

where E_0 is the projector onto the vacuum state and $dE(p)$ has no support for $p^2 < m^2$ for some $m > 0$. Choose a function $\tilde{g}(p_0)\in O_M(\mathbb{R}^1)$ which is real, even, and equals $|p_0|^{-1}$ for $|p_0| \geqslant m$. Define

$$\tau(B) = \int B(x^0)g(x^0)\,dx^0\,; \tag{2.20}$$

the integral exists and is an element of P_{QL} [cf. (2.22), (2.23) below for details]. τ is a linear functional, and $\tau(B)$ is Hermitian if B is, since g is real. Furthermore,

$$\begin{aligned} P_0\tau(B)\Omega &= P_0\left(\int U(x^0)g(x^0)dx^0\right)B\Omega \\ &= P_0\left(\int e^{ip_\mu x^\mu}g(x^0)dx^0\,dE(p)B\Omega\right) \\ &= \int p_0\tilde{g}(p_0)dE(p)B\Omega \\ &= \int dE(p)B\Omega = B\Omega - E_0B\Omega, \end{aligned}$$

by (2.19) for $x = 0$, as asserted.

(ii) By linearity, it is enough to prove (2.14) for $B = B^*\in P_{\mathrm{QL}}$:

$$\begin{aligned} &2(\Omega\,|[Q,B\,]\Omega) - (\Omega\,|[Q_r,B\,]\Omega) \\ &\quad = (B\Omega\,|Q_r\,\Omega) + (\Omega\,|Q_rB\Omega) \\ &\quad = (P_0\tau(B)\Omega\,|\,Q_r\Omega) + (\Omega\,|Q_rP_0\tau(B)\Omega) \\ &\quad = (\Omega\,|\tau(B)[P_0,Q_r\,]\Omega) - (\Omega\,|\,[P_0,Q_r\,]\tau(B)\Omega), \end{aligned}$$

since $P_0\Omega = 0$. $\square$

This lemma is sufficient to prove that $\Omega \times P_{\mathrm{QL}}\Omega$ is in the domain of Q and even controls the rate of convergence: Consider, along with the given current $J^\mu(x) = j^\mu(x,x)$ the quantity $\check{J}^\mu(x) = i[P_0,J^\mu(x)]$ which again is a current fulfilling (J1)–(J5). Thus, for the associated approximate charges $\check{Q}_r$ [cf. (1.5)] and functionals $\check{q}_r,\check{q}$ [cf. (2.6)], all previously obtained conclusions hold. By the last lemma, for $B\in P_{\mathrm{QL}}$ and $\varphi = \psi = \Omega$,

$$2(\Omega\,|Q_rB\Omega) = q_r(B) + i\check{q}_r(\tau(B)). \tag{2.21}$$

By Corollary 2.3, both terms on the right converge as $r\to\infty$, and

$$\begin{aligned} &2|Q(\Omega\,|B\Omega) - (\Omega\,|Q_rB\Omega)| \\ &\quad \leqslant |q_r(B) - q(B)| + |\check{q}_r(\tau(B)) - \check{q}(\tau(B))|\,. \end{aligned} \tag{2.22}$$

Now if $B = A(f)\in P_{\mathrm{QL}}$, then, by (2.20), $\tau(A(f)) = A(f^\tau)$ with

$$f^\tau(x_1,\ldots,x_l) := \int f(\mathbf{x}_1,x_1^0 - x^0,\ldots,\mathbf{x}_1,x_1^0 - x^0)g(x^0)\,dx^0\,. \tag{2.23}$$

However,

$$\|f^\tau\|_{m,n} \leqslant C\|f\|_{m,n}\,. \tag{2.24}$$

For the proof, note first that the space parts $\mathbf{x}_1,\ldots,\mathbf{x}_l$ in $\|\cdot\|_{m,n}$ are not affected by the replacement $f\to f^\tau$, so they might as well be suppressed, and so will the upper index in x_l^0 and x_0. Next suppose $l = 1$. Then

$$\begin{aligned} \|f^\tau\|_{s,t} &= \sup\left|\,|x|^s\left|\left(\frac{\partial}{\partial x}\right)^t f^\tau(x)\right|\,\right| \\ &\leqslant \sup\left[\int |x-y+y|^s\left|\left(\frac{\partial}{\partial(x-y)}\right)^t f(x-y)\right| \right. \\ &\qquad \left. \times|g(y)|dy\right] \\ &\leqslant \sum_{k=0}^{s} c_k \sup_{x,y}\left\{|x-y|^{s-k}\left|\left(\frac{\partial}{\partial(x-y)}\right)^t\right.\right. \\ &\qquad \left.\left. \times f(x-y)\right\}\cdot\int |y|^k|g(y)|\,dy, \end{aligned}$$

by the binomial formula. Consider first the supremum of $\{\cdot\}$ over x only, which is simply the translationally invariant norm $\|f\|_{s-k,t}$ and thus independent of y. Next, the integral is finite: Since $\tilde{g}\in L^2$ by construction, $g\in L^2$ and, in particular, locally integrable. Since g is infinitely differentiable, g is of rapid decrease, i.e.,

$$\|f^\tau\|_{s,t} \leqslant \sum_{k=0}^{s}\hat{C}_k\|f\|_{s-k,t}$$

which implies (2.24) since $\mathscr{S}$-norms are compatible. An analogous proof works for $l > 1$. All this can be summarized in the following

2.5 *Lemma*: For $A(f) \in P_{QL}$ there are constants k,n such that, for all positive integers N,

$$|Q(\Omega | A(f)\Omega) - (\Omega | Q_r A(f)\Omega)| \leqslant C_N (1+r)^{-N} \|f\|_{N+k,n} \tag{2.25}$$

with C_N independent of r.

Proof: Combine (2.22), (2.24), and Corollary 2.3. □

Finally, to show that Q is defined on $P_{QL}\Omega \times P_{QL}\Omega$, simply consider two operators $A(f)$, $B(g) \in P_{QL}$ and write

$$(A(f)\Omega | Q_r B(g)\Omega) = (A(f)\Omega | [Q_r, B(g)]\Omega) + (\Omega | (A(f))^* B(g) Q_r \Omega).$$

Corollary 2.3 and Lemma 2.5 imply that the right-hand side converges for $r \to \infty$ and even give the rate of convergence which proves

2.6 *Theorem*: For all $A(f), B(g) \in P_{QL}$,

$(A(f)\Omega | Q_r B(g)\Omega)$

converges, as $r \to \infty$, to a Hermitian sesquilinear form Q. Furthermore, there are constants k,n and, for all N, a constant C_N such that

$$|(A(f)\Omega | Q_r B(g)\Omega) - Q(A(f)\Omega | B(g)\Omega)| \leqslant C_N (1+r)^{-N} \|f\|_{N+k,n} \|g\|_{N+k,n}.$$

As a corollary, Q is continuous in the $\mathscr{S}$-seminorms:

2.7. *Corollary*: For all $A(f), B(g) \in P_{QL}$,

$$|Q(A(f)\Omega | B(g)\Omega)| \leqslant C \|f\|_{k,n} \|g\|_{k,n}$$

for some $\mathscr{S}(\mathbb{R}^4)$-norm $\| \ \|_{k,n}$.

Proof: By the triangle inequality,

$$|Q(A(f)\Omega | B(g)\Omega)| \leqslant |Q(A(f)\Omega | B(g)\Omega) - (A(f)\Omega | Q_r B(g)\Omega)| + |(A(f)\Omega | Q_r B(g)\Omega)|;$$

the second term is bounded by $(1+r)^k \|f\|_{l,m} \|g\|_{l',m'}$ since $(A(y)\Omega | j^0(x,x) B(z)\Omega)$ is a distribution in all variables, and the first term is bounded by $C_N \|f\|_{k,n} \|g\|_{k,n}$, by Theorem 2.6. Choose r fixed to get the result. □

3. SCATTERING STATES AND THE DOMAIN OF Q

It is the purpose of this section to show that Q can be defined on (a dense set of) scattering states and to determine its form there.

First, the Haag–Ruelle construction will be briefly reviewed. Assuming isolated one-particle mass hyperboloids of mass m_ν, one can find operators $B_\nu \in P_{QL}$ normalized so that [cf. (2.19)]:

$$(B_\nu \Omega | dE(p) B_\mu \Omega) = \delta(p^2 - m_\nu^2) \delta_{\nu\mu} \sigma_\nu(\mathbf{p}) d^4 p$$

with a strictly positive, infinitely often differentiable function σ_ν. With the help of smooth solutions of the Klein–Gordon equation for mass $m > 0$,

$$f(\mathbf{x},t) := \int \exp\{-i[\omega(m,\mathbf{p})t - \mathbf{px}]\} \check{f}(\mathbf{p}) \times [\omega(m,\mathbf{p})]^{-1} d^3 p, \tag{3.1}$$

$\check{f} \in \mathscr{S}(\mathbb{R}^3)$, $[\omega(m,\mathbf{p})]^2 := \mathbf{p}^2 + m^2$,

define operators

$$B(f,t) := \int U(\mathbf{a},t) B U(-\mathbf{a},t) f(\mathbf{a},t) d^3 a. \tag{3.2}$$

Then the state

$$\varphi(t) \equiv \phi(t)\Omega \equiv \varphi_{\nu_1 \cdots \nu_n}(f_1,\ldots,f_n;t) := B_{\nu_1}(f_1,t) \cdot \cdots \cdot B_{\nu_n}(f_n,t)\Omega \tag{3.3}$$

converges strongly, for $t \to \pm\infty$, to some limit, denoted by

$$\varphi^{\text{in}} \equiv \varphi^{\text{in}}_{\nu_1 \cdots \nu_n}(\hat{f}_1,\ldots,\hat{f}_n). \tag{3.4}$$

This state is interpreted as an outgoing or incoming free n-particle state characterized by quantum numbers ν (in particular masses m_ν) and wave functions $\hat{f}_\nu$ [assumed normalized with respect to $d\omega(m_\nu,\mathbf{p})$], where

$\hat{f}_\nu(\mathbf{p}) := \check{f}_\nu(\mathbf{p}) [\sigma_\nu(\mathbf{p})]^{1/2}$.

The reason is that

$$(a_\nu^{\text{in}}(\hat{f}_\nu))^* \varphi^{\text{in}}_{\nu_1 \cdots \nu_n}(\hat{f}_1,\ldots,\hat{f}_n) = \lim_{t \to \pm\infty} B_\nu(f_\nu,t) B_{\nu_1}(f_1,t) \cdot \cdots \cdot B_{\nu_n}(f_n,t)\Omega \tag{3.5}$$

defines a free field creation operator

$$(a_\nu^{\text{ex}}(\hat{f}_\nu))^* = f(a_\nu^{\text{ex}})^*(\mathbf{p}) \hat{f}_\nu(\mathbf{p}) d\omega(m_\nu,\mathbf{p})$$

with adjoint $a^{\text{ex}}{}_\nu$ ("ex" stands for either "in" or "out") satisfying the usual commutation relations

$$[a_\mu^{\text{ex}}(\mathbf{p}), (a_\nu^{\text{ex}})^*(\mathbf{q})]_\pm = 2\delta_{\nu\mu} \delta(\mathbf{p} - \mathbf{q}) \omega(m_\nu,\mathbf{p}), \tag{3.6}$$

such that

$$\varphi^{\text{ex}}_{\nu_1 \cdots \nu_n}(\hat{f}_1,\ldots,\hat{f}_n) = (a_{\nu_1}^{\text{ex}}(\hat{f}_1))^* \cdot \cdots \cdot (a_{\nu_n}^{\text{ex}}(\hat{f}_{\nu_n}))^* \Omega. \tag{3.7}$$

The Fock space constructed from $a_\mu^{\text{ex}}, (a_\mu^{\text{ex}})^*$ will be denoted by H^{ex}.

There are assertions on the rate of convergence in (3.3), (3.4). Define the velocity support $\Sigma(f)$ of a smooth solution of a Klein–Gordon equation with mass m by

$$\sum(f) := \{\mathbf{p} \cdot (p^2 + m^2)^{-1/2} | \mathbf{p} \in \operatorname{supp} \check{f}\}$$

and denote the space generated by all states $\varphi^{\text{ex}}_{\nu_1 \cdots \nu_n}(\hat{f}_1,\ldots,\hat{f}_n)$ such that the $\Sigma(f_i)$ are pairwise disjoint by D_n^{ex}. For $\varphi^{\text{ex}} \in D^{\text{ex}}$, the space generated by $\cup_n D_n^{\text{ex}}$,

$$\|\varphi(t) - \varphi^{\text{ex}}\| \leqslant A_N (1+t)^{-N} \tag{3.8}$$

for all N, where $\varphi(t)$ is a finite linear combination of states of the form (3.3). $D^{\text{ex}} \cup \{\Omega\}$ is dense in H^{ex}. This concludes the results that will be needed in the following.

It is in considering Q on scattering states tht one sees most clearly the difference of the present approach to the one in Refs. 3 and 4, where unbroken symmetries were treated.

If the symmetry is not spontaneously broken, Q, defined in (1.8), is a symmetric operator, hence closable. To define Q on scattering states, it is thus enough to prove that $Q\varphi_t$ converges in t. Once one knows this, and since one can also show that Q is the quadratic form limit of Q_r, Q as a form on scattering states ϕ, ψ approximated by $\varphi(t)$, $\psi(t)$ is

given by

$$(\psi|Q\varphi) = \lim_{t\to\pm\infty}\lim_{r\to\infty}(\psi(t)|Q_r\,\varphi(t)), \tag{3.9}$$

in that order; this was done in Refs. 3 and 4.

On the other hand, if the symmetry is spontaneously broken, Q, defined in (1.8), is not closable. For that reason, the quadratic form Q was defined from the outset as the limit of the form Q_r. To show that Q is defined on scattering states, one has now to prove

$$Q(\psi|\varphi) = \lim_{r\to\infty}(\psi|Q_r\,\varphi) = \lim_{r\to\infty}\lim_{r\to\pm\infty}(\psi(t)|Q_r\,\varphi(t)), \tag{3.10}$$

where the repeated limits are now in reversed order.

The results of Ref. 3 lead one to suspect that even in the case of spontaneously broken symmetries, the limit (3.9) is easier to handle than (3.10). So the next task will be to prove that the limits (3.9) and (3.10) are equal also in the present case.

This will follow from the estimates (valid for all N)

$$|(\psi|Q_r\varphi) - (\psi(t)|Q_r\varphi(t))| \leqslant A_N(1+r)^K(1+|t|)^{-N} \tag{3.11}$$

$$|Q(\psi(t)|\varphi(t)) - (\psi(t)|Q_r\varphi(t))| \leqslant B_N(1+r)^{-N}(1+|t|)^{cN} \tag{3.12}$$

where $\varphi,\psi\in D^{\rm ex}$ [approximated by $\varphi(t),\psi(t)$ as defined in (3.3)] and K,c are constants independent of N. To prove this, observe first that

$$\|Q_r\varphi(t)\|^2 \leqslant P(r,t), \tag{3.13}$$

where P is a polynomial in r and t; this follows simply by considering

$$(B_{\nu_1}(\mathbf{a}_1,t)\cdot\cdots\cdot B_{\nu_n}(\mathbf{a}_n,t)\Omega\,|\,j^0(x,x) \times j^0(y,y)B_{\nu_1}(\mathbf{b}_1,t)\cdot\cdots\cdot B_{\nu_n}(\mathbf{b}_n,t)\Omega),$$

which is polynomially bounded in $\mathbf{a}_\nu,\mathbf{b}_\mu$, t and a distribution in x and y, by the distribution properties of Wightman functions. Integrating over $\mathbf{a}_\nu$ and $\mathbf{b}_\nu$ with f_ν and over x,y with $\vartheta_r\otimes\eta$ gives the bound

$$\|Q_r\varphi(t)\|^2 \leqslant C_{kl}\left(\int P_1(\mathbf{a}_1,\ldots,\mathbf{a}_{n,t}t) \prod_{\nu=1}^{n}|f_\nu(\mathbf{a}_\nu,t)|d^3a_1\cdots d^3a_n\right)^2\cdot\|\vartheta_r\otimes\eta\|_{k,l}^2,$$

where P_1 is some polynomial. But the $\|\ \|_{k,l}$-norm of $\vartheta_r\otimes\eta$ is bounded by a polynomial in r, and the integral by a polynomial in t (this follows from well-known properties of smooth Klein–Gordon solutions, see, e.g., Lemma 6.3 of Ref. 3). Thus, (3.13) is proved.

To establish (3.11), observe that the left-hand side is bounded by

$$|(\psi-\psi(t)|Q_r\varphi)| + |(Q_r\psi(t)|\varphi-\varphi(t))| \leqslant \|\psi-\psi(t)\|\,\|Q_r\varphi\| + \|\varphi-\varphi(t)\|\,\|Q_r\psi(t)\|$$

so that (3.13) and (3.8) imply the inequality.

As for (3.12), note first that $\varphi(t)$ is, for fixed t, in $P_{\rm QL}\Omega$: Assume that in (3.2), $B\in P_{\rm SL}$ has the representation

$$B = \int A_1(x_1)\cdot\cdots\cdot A_l(x_l)g(x_1\cdots x_l)d^4x_1\cdots d^4x_l$$

Then $B(f,t)$ has an analogous representation with g replaced by

$$\int g(x_1-(\mathbf{a},t),\ldots,x_l-(\mathbf{a},t)f(\mathbf{a},t)d^3a \equiv g_t(x_1\cdots x_l).$$

But any $\mathscr{S}(\mathbb{R}^{4l})$-norm of g_t is polynomially bounded in t: As in the proof of (2.24),

$$\|g_t\|_{a,b} \leqslant \sum_{k=0}^{a} C_k\|g\|_{a-k,b}\int|(\mathbf{b},t)|^k\,|f(\mathbf{b},t)|d^3b, \tag{3.14}$$

where the integral is bounded by $(1+|t|)^{2k+3/2}$ (see again Lemma 6.2 of Ref. 3). Now Theorem 2.6 implies (3.12).

Next, suppose $Q(\psi(t),\varphi(t))$ tends to a limit as $t\to\infty$; call that limit $\widehat{Q}(\psi|\varphi)$. To show that $(\psi|Q_r\varphi)$ tends to the same limit, note that by (3.11) and (3.12),

$$\begin{aligned}|\widehat{Q}(\psi|\varphi) - (\psi|Q_r\varphi)| &\leqslant |\widehat{Q}(\psi|\varphi) - Q(\psi(t)|\varphi(t))| + |Q(\psi(t)|\varphi(t)) - (\psi(t)|Q_r\varphi(t))| + |(\psi(t)|Q_r\varphi(t)) - (\psi|Q_r\varphi)| \\ &\leqslant |\widehat{Q}(\psi|\varphi) - Q(\psi(t)|\varphi(t))| + B_N(1+r)^{-N}(1+|t|)^{cN} + A_N(1+r)^K(1+|t|)^{-N};\end{aligned}$$

choose $N>K$ and $|t| = r^{1/(c+1)}$. For r big enough, the first term is small by assumption, while the last two terms are small since they are bounded by $C_N(1+r)^{K-N}$. This shows that $\widehat{Q}=Q$.

To investigate the limit (3.9), the sesquilinear form $Q(\psi(t)|\varphi(t))$ will be expanded in truncated vacuum expectation values (denoted by $\langle\ \rangle^T$). Suppose [cf. (3.3)]

$$\psi(t) = \Psi(t)\Omega, \quad \varphi(t) = \Phi(t)\Omega \tag{3.15}$$

and define

$$\{Q(\psi(t)|\varphi(t))\}^T := \lim_{r\to\infty}\langle\Psi^*(t)Q_r\Phi(t)\rangle^T. \tag{3.16}$$

Since $\langle\Psi^*(t)Q_r\Phi(t)\rangle^T$ is obtained by a finite number of algebraic operations from $(\Omega\,|\,\Psi^*(t)Q_r\Phi(t)\Omega)$, the results of Sec. 2 imply the existence of the r limit. Furthermore, the following generalization of Lemma 7.1 in Ref. 3 holds which will be proven in the Appendix.

3.1 *Lemma*: Let $\varphi\in D_n^{\rm ex},\psi\in D_m^{\rm ex}$ be approximated by vectors $\varphi(t),\psi(t)$ of the form (3.3), and suppose $n\geqslant 2$ or $m\geqslant 2$. Then for all N,

$$|\{Q(\psi(t)|\varphi(t))\}^T| \leqslant C_N(1+|t|)^{-N}. \tag{3.17}$$

This leads immediately to

3.2 *Theorem*: Let $\varphi,\psi\in D^{\rm ex}$ be approximated by vectors $\varphi(t),\psi(t)$ of the form (3.3). Then

$$\begin{aligned}Q(\psi|\varphi) = \sum_{i,k}\int &(\psi|\,(a_i^{\rm ex})^*(\mathbf{q})a_k^{\rm ex}(\mathbf{p})\varphi) \\ &\cdot Q((a_i^{\rm ex})^*(\mathbf{q})\Omega\,|\,(a_k^{\rm ex})^*(\mathbf{p})\Omega)d\omega(m_i,\mathbf{q}) \\ &\times d\omega(m_k,\mathbf{p}) + \sum_i\int(\psi|(a_i^{\rm ex})^*(\mathbf{q})\varphi) \\ &\cdot Q((a_i^{\rm ex}(\mathbf{q}))^*\Omega\,|\,\Omega)d\omega(m_i,\mathbf{q}) \\ &+ \sum_k\int(\psi|a_k^{\rm ex}(\mathbf{p})\varphi) \\ &\times Q(\Omega\,|\,(a_k^{\rm ex}(\mathbf{p}))^*\Omega)d\omega(m_k,\mathbf{p}).\end{aligned} \tag{3.18}$$

Note that the first term, bilinear in $(a^{\mathrm{ex}})^*$, is the one that appears in the case of an unbroken symmetry. The additional terms, linear in a^{ex}, $(a^{\mathrm{ex}})^*$ are the ones responsible for the spontaneous symmetry breakdown.

Proof: By the discussion before Lemma 7.1,

$$\begin{aligned} Q(\psi|\varphi) &= \lim_{t\to\pm\infty} Q(\psi(t)|\varphi(t)) \\ &= \lim_{t\to\pm\infty}\lim_{r\to\infty}(\psi(t)\mid Q_r\varphi(t)) \\ &= \lim_{t\to\pm\infty}\lim_{r\to\infty}\int(\psi(t)|j^0(x,x)\varphi(t)) \\ &\quad\times\vartheta_r(\mathbf{x})\eta(x^0)d^4x. \end{aligned}$$

Following Araki and Haag,[7] the expansion of the integrand in truncated vacuum expectation values yields five types of expressions:

(i) $\langle j^0\rangle^T\cdot\left\{\sum\prod\langle\cdots\rangle^T\right\}$,

(ii) $\langle j^0 B_{\mu_k}(g_k,(t)\rangle^T\left\{\sum\prod\langle\cdots\rangle^T\right\}$,

(iii) $\langle(B_{\nu_i}(f_i,t))^* j_0\rangle^T\cdot\left\{\sum\prod\langle\cdots\rangle^T\right\}$,

(iv) $\langle(B_{\nu_i}(f_i,t))^* j_0 B_{\mu_k}(g_k,t)\rangle^T\cdot\left\{\sum\prod\langle\cdots\rangle^T\right\}$,

(v) $\left\langle\prod_{l=1}^{m}(B_{\nu_l}(f_l,t))^* j_0\prod_{s=1}^{n}B_{\mu_s}(g_s,t)\right\rangle^T\cdot\left\{\sum\prod\langle\cdots\rangle^T\right\}$,

with at least two factors either before or after j^0. (i) is zero by (j5), and the limit, as $r\to\infty$, of (v) vanishes in the limit $t\to\infty$ by Lemma 3.1. The first factors in (ii), (iii), and (v) are all time independent and thus given by

$$\begin{aligned} &\langle j^0(a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k))^*\rangle^T,\quad \langle a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i)j^0\rangle^T, \\ &\langle a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i)j^0(a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k))^*\rangle^T, \end{aligned} \tag{3.19}$$

where the superscript T can be dropped: The expansion of the truncated vacuum expectation values in untruncated ones always contains a factor $(\Omega\,|a^{\mathrm{ex}}\Omega)=0$. The brackets $\{\cdots\}$ in (ii), (iii), and (iv) are independent of the integration variable x and thus independent of r. Integrating (3.19) yields in the limit $r\to\infty$,

$$\begin{aligned} &Q(\Omega\,|(a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k))^*\Omega),\quad Q((a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i))^*\Omega\,|\Omega), \\ &Q((a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i))^*\Omega\,|(a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k))^*\Omega), \end{aligned}$$

where the limits exist by Lemma 2.5 since $(a_{\nu_i}(\hat{f}_i))^*\Omega \equiv B_{\nu_i}(f_i,t)\Omega$ is in $P_{\mathrm{QL}}\Omega$. The t-dependent brackets $\{\cdots\}$ in (ii), (iii), and (iv) are, in the limit $t\to\pm\infty$, just expansions in truncated vacuum expectation values of $(\psi|\varphi_k)$,$(\psi_i|\varphi)$, $(\psi_i|\varphi_k)$ where φ_k, resp. ψ_i, is obtained from φ, resp. ψ, by deleting $(a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k))^*$ resp. $(a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i))^*$. Note that, e.g.,

$$\varphi_k=\epsilon_k a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k)\varphi,$$

with $\epsilon_k=\pm 1$ depending on the number of anticommutations required by (3.6) (recall that the $\hat{g}_k$ have disjoint support and are normalized), and similarly for ψ_i. Hence the brackets converge to

$$\begin{aligned} &\epsilon_k(\psi|a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k)\varphi),\quad \epsilon_i(\psi|(a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i))^*\varphi), \\ &\epsilon_i\epsilon_k(\psi|(a_{\nu_i}{}^{\mathrm{ex}}(\hat{f}_i))^* a_{\mu_k}{}^{\mathrm{ex}}(g_k)\varphi). \end{aligned}$$

But

$$\begin{aligned} &\sum_k\epsilon_k(\psi|a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k)\varphi)\cdot Q(\Omega\,|(a_{\mu_k}{}^{\mathrm{ex}}(\hat{g}_k))^*\Omega) \\ &\quad=\sum_l\int(\psi|a_l{}^{\mathrm{ex}}(\mathbf{q})\varphi)\cdot Q(\Omega\,|(a_l{}^{\mathrm{ex}}(\mathbf{q}))^*\Omega)d\omega(m_l,\mathbf{q}) \end{aligned}$$

by (3.6), and similarly for the other terms. □

Note that, by previous results, (3.18) remains true if $\varphi=\Omega$ or $\psi=\Omega$.

Thus, Q is a sum of a term bilinear in creation and annihilation operators and two linear terms; hence Q is zero between states of D^{ex} whose particle number differs by more than one. If the symmetry is not spontaneously broken, the linear terms are absent by Lemma 2.4 and Q commutes with the number operator $\hat{N}$ (in the sense of forms) i.e.,

$$Q(\psi|\hat{N}\varphi)=Q(\hat{N}\psi|\varphi). \tag{3.20}$$

This coincides with the results of Ref. 2. The converse is also true under additional assumptions; (3.20) implies that the linear terms in (3.18) are absent for all $\varphi,\psi\in D^{\mathrm{ex}}\cup\Omega$. Choose $\psi=(a_\nu{}^{\mathrm{ex}}(g))^*\Omega$, $\varphi=\Omega$ to obtain

$$\begin{aligned} 0=Q(\Omega\,|(a_\nu{}^{\mathrm{ex}}(g))^*\Omega) &= \lim_{r\to\infty}(\Omega\,|Q_r(a_\nu{}^{\mathrm{ex}}(g))^*\Omega) \\ &= \lim_{r\to\infty}(\Omega\,|Q_r\phi_\nu{}^{\mathrm{ex}}(g)\Omega), \end{aligned}$$

where $\phi_\nu{}^{\mathrm{ex}}$ are the free asymptotic fields constructed from $a_\nu{}^{\mathrm{ex}}$. Assume asymptotic completeness and the existence of interpolating fields so that

$$\phi_\nu(g)\Omega=\phi_\nu{}^{\mathrm{ex}}(g)\Omega,\quad \lim_{r\to\infty}(\Omega\,|Q_r\phi_\nu(g)\Omega)=0.$$

Take g real:

$$\lim_{r\to\infty}(\Omega\,|[Q_r,\phi_\nu(g)]\Omega)=2i\,\mathrm{Im}\lim_{r\to\infty}(\Omega\,|Q_r\phi_\nu(g)\Omega)=0,$$

where "Im" denotes the imaginary part. Thus, (1.14) holds for ϕ_ν and hence for P_{SL}. In short, the symmetry is unbroken if and only if it does not change the particle number, and it is spontaneously broken if and only if it has a nonvanishing component changing the particle number by one.

Note further that the integral kernels in (3.18) depend only on one-particle states which are the same for "in" and "out" labels. Hence, Q commutes with the S matrix:

3.3 *Theorem*: For $\psi^{\mathrm{in}}\in D^{\mathrm{in}},\varphi^{\mathrm{out}}\in D^{\mathrm{out}}$ such that $S^*\psi^{\mathrm{in}}\in D^{\mathrm{out}},S\varphi^{\mathrm{out}}\in D^{\mathrm{in}}$,

$$Q(\psi^{\mathrm{in}}|S\varphi^{\mathrm{out}})=Q(S^*\psi^{\mathrm{in}}|\varphi^{\mathrm{out}}). \tag{3.21}$$

Proof: By the unitary of the S matrix

$$(\psi^{\mathrm{in}}|Sa_\kappa^{\mathrm{out}}\varphi^{\mathrm{out}})=(S^*\psi^{\mathrm{in}}|a_\kappa^{\mathrm{out}}\varphi^{\mathrm{out}}),$$

and similarly for the other scalar products in (3.18). Multiply with the "ex"-independent kernels Q and integrate to obtain (3.21). □

4. CONCLUSIONS

The general form of spontaneously broken symmetries arising from translationally noncovariant currents in theor-

ies with mass gap has been determined. The symmetry generator exists as a form on a dense set of asymptotic states, commutes with the S matrix, and has the form given by Theorem 3.2. The results are incomplete in several respects: No conclusions can be drawn from them for spontaneous symmetry breaking in the case of no mass gap nor in the case of covariant currents (for which Goldstone's theorem then precludes the existence of a mass gap). Furthermore, restrictions on the integral kernels in Theorem 3.2 must be expected to be similar to those obtained in Ref. 3. However, for the present case it is not trivial to determine under what conditions $[Q,\phi_v^{ex}(x)]$ [where $\phi_v^{ex}(x)$ is the free field constructed from a_v^{ex}], is localized in x. For the examples of free fields in the Introduction this is certainly true so that, by the methods of Ref. 3, Q is a superposition of examples (2′) (with more general polynomials) and (4′).

ACKNOWLEDGMENT

I would like to thank the members of the University of Princeton for their kind hospitality during the academic year 1977/78 when this work was completed. Special thanks are due to Barry Simon for supplying a proof of Lemma A2 which is certainly simpler than my original abstract one; he was, in his own words, spurred into action by his belief that "the only purpose of general abstract nonsense is to keep people off the street." Finally, without the constant encouragement of Helmut Reeh this work would never have been undertaken.

APPENDIX

The purpose of this Appendix is to prove Lemma 3.1. It will be convenient to introduce, along with the given current $J^\mu(x)=j^\mu(x,x)$, the related current density

$$j_f^\mu(y,x):=\int j^\mu(y,x+z)f(z)d^4z, \tag{A1}$$

where $f\in\mathcal{S}(\mathbb{R}^4)$ is real. Then $j_f^\mu(y,x)$ satisfies (j0)–(j5), giving rise to operators Q_t^f and the sesquilinear form Q^f. Also:

$$j_f^\mu(x):=j_f^\mu(x,x)=\int \tilde{j}^\mu(x,p)e^{ipx}\tilde{f}(p)\,d^4p, \tag{A2}$$

where $\tilde{j}^\mu(x,p)$ denotes the Fourier transform of $j^\mu(x,z)$ in z. The first step will be to prove Lemma 3.1 for Q^f.

A 1. Lemma: Suppose $\varphi\in D_1^{ex}, \psi\in D_n^{ex}$ are approximated by vectors $\varphi(t)$, $\psi(t)$ of the form (3.3), and assume $1\geqslant 2$ or $n\geqslant 2$. Then there is a $\mathcal{S}(\mathbb{R}^4)$-norm $\|\ \|_{r,s}$ such that for all N and $f\in\mathcal{S}(\mathbb{R}^4)$

$$|\{Q^f(\psi(t)|\varphi(t))\}^T|\leqslant C_N(1+|t|)^{-N}\|\tilde{f}\|_{r,s}. \tag{A3}$$

This lemma is a slightly generalized version of Lemma 7.1 in Ref. 3; the f-dependence in the estimate is now explicit. The proof will follow along the lines in Ref. 3 and requires some preliminary lemmas.

Assume the vectors $\varphi(t)$ and $\psi(t)$ to be of the form

$$\varphi(t):=\int \Phi(\mathbf{a}_1,\ldots,\mathbf{a}_l,t)f_1(\mathbf{a}_1,t)\cdot\cdots\cdot f_l(\mathbf{a}_l,t)d^3a_1\cdots d^3a_l\,\Omega,$$

$$\psi(t):=\int \Psi(\mathbf{a}_{l+1},\ldots,\mathbf{a}_{l+n},t)f_{l+1}(\mathbf{a}_{l+1},t)\cdot\cdots \tag{A4}$$

$$\times\cdot f_{l+n}(\mathbf{a}_{l+n},t)d^3a_{l+1}\cdots d^3a_{l+n}\Omega,$$

where

$$\Phi(\mathbf{a}_1,\ldots,\mathbf{a}_l,t):=B_1(\mathbf{a}_1,t)\cdot\cdots\cdot B_l(\mathbf{a}_l,t),$$

$$\Psi(\mathbf{a}_{l+1},\ldots,\mathbf{a}_{l+n},t):=B_{l+1}(\mathbf{a}_{l+1},t)\cdot\cdots\cdot B_{l+n}(\mathbf{a}_{l+1},t), \tag{A5}$$

with $B_i\in P_{QL}$. Consider the truncated vacuum expectation value

$$W_t(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};x,p):=\langle\Psi^*\tilde{j}^0(x,p)\Phi\rangle^T \tag{A6}$$

and note that

$$\hat{W}_t(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};\tilde{f}):$$

$$=\int W_t(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};x,p)e^{ipx}\tilde{f}(p)\eta(x^0)d^4pd^4x \tag{A7}$$

is related to $\{Q^f(\psi(t)|\varphi(t))\}^T$ by

$$\{Q^f(\psi(t)|\varphi(t))\}^T$$

$$=\int \hat{W}_t(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};\tilde{f})\prod_{i=1}^n f_i(\mathbf{a}_i,t)d^3a_1\cdots d^3a_{l+n}, \tag{A8}$$

where f_i are the smooth solutions of Klein–Gordon equations from (A4). Thus, the study of $\{Q^f\}^T$ is reduced to $\hat{W}_t$ and to W_t; the latter will turn out to be a test function from $\mathcal{S}$ in some variables and a tempered distribution in others. Such objects obey the following estimate:

A 2. Lemma: For $T(x,y)\in\mathcal{S}(\mathbb{R}^4)\otimes\mathcal{S}'(\mathbb{R}^l)$, there is a $\mathcal{S}$-norm $\|\ \|_{r,s}$ such that for all n and $f\in\mathcal{S}(\mathbb{R}^l)$,

$$|T(x,f)|\leqslant C_n(1+|x|^2)^{-n}\|f\|_{r,s}, \tag{A9}$$

where C_n does not depend on x and f.

Proof: Since $T\in\mathcal{S}(\mathbb{R}^k)\otimes\mathcal{S}'(\mathbb{R}^l)$,

$$|T(x,f)|\leqslant C_x\|f\|_{r,s}, \tag{A10}$$

$$|T(x,f)|\leqslant C_f^n(1+|x|^2)^{-n}. \tag{A11}$$

Fix a positive integer n. By (A10)

$$S_x(f):=(1+|x|^2)^nT(x,f)$$

is, for fixed x, a continuous functional on $\mathcal{S}(\mathbb{R}^l)$. By (A11), the family of functionals $\{S_x\,|x\in\mathbb{R}^k\}$ is pointwise bounded:

$$|S_x(f)|\leqslant C_f.$$

The uniform boundedness principle for the Fréchet space $\mathcal{S}(\mathbb{R}^l)$ now implies (Ref. 8, p. 172)

$$|S_x(f)|\leqslant C \tag{A12}$$

with C independent of x and f if f varies over a bounded set in $\mathcal{S}(\mathbb{R}^l)$, say over the bounded set $\{f\in\mathcal{S}(\mathbb{R}^l)|\|f\|_{r,r}<1\}$. Replace f by $f(\|f\|_{r,r})^{-1}$ in (A12) to get (A9). □

The last lemma implies the following estimate on $\hat{W}_t$.

A 3. Lemma: There is a $\mathcal{S}(\mathbb{R}^4)$-norm $\|\ \|_{r,s}$ such that, for all N,

$$|\hat{W}_t(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};\tilde{f})|\leqslant C_n\|\tilde{f}\|_{r,s}(1+|t|)^k(1+|\mathbf{a}_l|^2)^k$$

$$\times\prod_{i=1}^{l+n-1}(1+|\mathbf{a}_i-\mathbf{a}_{i+1}|^2)^{-N}. \tag{A13}$$

Proof: Translation covariance and current conservation imply

$$\hat{W}_t(\mathbf{a}_1,\ldots,\mathbf{a}_{1+n};\tilde{f})=\hat{W}_0(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};\tilde{f}_t) \tag{A14}$$

with

$$\tilde{f}_t(p) := e^{ip_0 t}\tilde{f}(p).$$

To bound $\hat{W}_0$, use (A7) and note that $W_0(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};x,p)$ is a test function from $\mathscr{S}$ in the difference variables

$$\mathbf{a}_1-\mathbf{a}_2,\ldots,\mathbf{a}_l-\mathbf{x},\mathbf{x}-\mathbf{a}_{l+1},\ldots,\mathbf{a}_{l+n-1}-\mathbf{a}_{l+n}$$

and a tempered distribution in x and p. By Lemma A2,

$$|W_0(\mathbf{a}_1,\ldots,\mathbf{a}_{l+n};\mathbf{x},g(x^0,p)|$$

$$\leqslant C_N\|g\|_{r,s}\prod_{i\neq l}(1+|\mathbf{a}_i-\mathbf{a}_{i+1}|^2)^{-N}\times(1+|\mathbf{a}_l-\mathbf{x}|^2)^{-N}(1+|\mathbf{x}-\mathbf{a}_l|^2)^{-N}. \tag{A15}$$

Now choose

$$g(x^0,p) := \exp[ip^0(x_0+t)]\exp(-i\mathbf{pb})\tilde{f}(p)\eta(x^0) \tag{A16}$$

for some $\mathbf{b}\in\mathbb{R}^3$ so that

$$\|g\|_{r',s'}\leqslant P(|\mathbf{b}|,t)\|f\|_{r,s} \tag{A17}$$

for some $\mathscr{S}$-norm $\|\ \|_{r,s}$ and some polynomial P of degree K, say. Insert (A16) in (A15), use (A17) and put $\mathbf{b}=\mathbf{x}$. Integrating the resulting inequality over $\mathbf{x}$ then proves (A14), since the $\mathbf{x}$ integral on the right,

$$I := \int(1+|\mathbf{x}|^2)^K(1+|\mathbf{a}_l-\mathbf{x}|^2)^{-N}\times(1+|\mathbf{x}-\mathbf{a}_{l+1}|^2)^{-N}d^3x, \tag{A18}$$

can be bounded as follows ($\boldsymbol{\alpha} := \mathbf{a}_1-\mathbf{a}_{l+1}$):

$$|I|\leqslant\int(1+|\mathbf{y}+\mathbf{a}_l|^2)^K(1+|\mathbf{y}|^2)^{-N}(1+|\mathbf{y}-\boldsymbol{\alpha}|^2)^{-N}d^3y$$

$$\leqslant C_K(1+|\mathbf{a}_l|^2)^K(1+|\boldsymbol{\alpha}|^2)^{K-N}\int(1+|\mathbf{y}|^2)^{K-N}d^3y \tag{A19}$$

where the inequality

$$1+|\mathbf{a}+\mathbf{b}|^2\leqslant 2(1+|\mathbf{a}|^2)(1+|\mathbf{b}|^2)$$

has been used for the first factor in (A19) with $\mathbf{a}=\mathbf{y}$ and $\mathbf{b}=\mathbf{a}_1$, and for the last two factors with $\mathbf{a}=\mathbf{y}$ and $\mathbf{b}=\boldsymbol{\alpha}-\mathbf{y}$. □

Proof of Lemma A 1: Integrate (A13) over $\mathbf{a}_1,\ldots,\mathbf{a}_{l+n}$ with smooth Klein–Gordon solutions $f_i(\mathbf{a}_i,t)$ and use Lemma 6.4 of Ref. 3. □

The next step will be to prove Lemma A1 with Q^f replaced by Q. Note that by (A1)

$$Q^f(A\Omega\,|B\Omega)$$

$$=\lim_{r\to\infty}\int(A\Omega\,|\,j^0(x,x+z)B\Omega)\vartheta_r(\mathbf{x})\eta(x^0)f(z)d^4xd^4z$$

$$=\lim_{r\to\infty}\int(A\Omega\,|\,j^0(x-z,x)B\Omega)\vartheta_r(\mathbf{x})\eta(x^0)f(z)d^4xd^4z$$

$$=\int Q(A(z)\Omega\,|B(z)\Omega)f(z)d^4z. \tag{A20}$$

Now consider a δ-sequence in $\mathbb{R}^4$,

$$f_n(z) := n^4f(nz) \tag{A21}$$

with

$$f\in\mathscr{D}(\mathbb{R}^4),\quad 0\leqslant f\leqslant 1,\quad \operatorname{supp}f\subset\{z\in\mathbb{R}^4|\ |z|\leqslant 1\},$$

$$\int f(z)d^4z=1.$$

Then, by Corollary 2.7 and (A20),

$$|Q^{f_n}(A(h)\Omega\,|B(g)\Omega)-Q(A(h)\Omega\,|B(g)\Omega)|$$

$$\leqslant C'\cdot\sup_{|z|\leqslant n^{-1}}\{\|h_z-h\|_{k,l}\|g_z\|_{k,l}+\|h\|_{k,l}\|g_z-g\|_{k,l}\} \tag{A22}$$

$$\leqslant C''n^{-1}\|h\|_{k',l'}\|g\|_{k',l'}$$

since

$$\|g_z\|_{k,l}\leqslant\hat{C}(1+|z|^2)^k\|g\|_{k,l},\quad \|g-g_z\|_{k,l}\leqslant|z|\|g\|_{k,l+1}.$$

Substitute $\varphi(t)$, resp. $\psi(t)$, for $A(h)$, Ω, resp. $B(g)\Omega$; the norms on the right of (A22) then grow polynomially in t so that

$$|Q^{f_n}(\psi(t)|\varphi(t))-Q(\psi(t)|\varphi(t))|\leqslant Cn^{-1}(1+|t|)^k \tag{A23}$$

and a similar equation is valid for the corresponding truncated quantities which are finite algebraic combinations of expressions (A23),

$$|\{Q^{f_n}(\psi(t)|\varphi(t))\}^T-\{Q(\psi(t)|\varphi(t))\}^T|\leqslant Cn^{-1}(1+|t|)^k. \tag{A24}$$

Finally, by Lemma A1,

$$|\{Q^{f_n}(\psi(t)|\varphi(t))\}^T|\leqslant C_N n^{2N}(1+|t|)^{-N}. \tag{A25}$$

Combining the last two equations gives

$$|\{Q(\psi(t)|\varphi(t))\}^T|\leqslant Cn^{-1}(1+|t|)^k+C_N n^{2N}(1+|t|)^{-N}. \tag{A26}$$

Fix a positive integer M, choose $N=2(k+M)s+M$ in (A26) and put $n=|t|^\alpha$ with $\alpha=k+M$ to get

$$|\{Q(\psi(t)|\varphi(t))\}^T|\leqslant C_M(1+|t|)^{-M}$$

which proves Lemma 3.1. □

[1]Y. Dothan and E.Gal-Ezer, Nuovo Cimento A **12**, 465–79 (1972).
[2]K. Kraus and L.J. Landau, Commun. Math. Phys. **24**, 243–52 (1972).
[3]W.D. Garber and H. Reeh, J. Math. Phys. **19**, 59–66 (1978).
[4]W.D. Garber and H. Reeh, J. Math. Phys. **19**, 985–6 (1978).
[5]R. Ferrari, Nuovo Cimento A **14**, 386–402 (1973).
[6]D. Kastler, D.W. Robinson, and A. Swieca, Commun. Math. Phys. **2**, 108–20 (1966).
[7]H. Araki and R. Haag, Commun. Math. Phys. **4**, 77–91 (1967).
[8]G. Köthe, *Topologische Lineare Räume* (Springer, Berlin, Heidelberg, New York, 1960).

Classification of gauge fields

Y. M. Cho
Max-Planck-Institut für Physik und Astrophysik, München, West Germany

(Received 17 April 1979; accepted for publication 6 August 1979)

A recently proposed classification of non-Abelian gauge fields is reexamined and refined to accommodate the internal symmetry structure of an arbitrary group. The concept of the canonical forms and the canonical frames is introduced and utilized to represent all the physically inequivalent realizations for each type of the field strengths. The local stability of the classification is shown to be minimal.

I. INTRODUCTION

In general relativity the algebraic classification of gravitational fields by Petrov[1] has been useful to provide us a deeper insight to the theory of gravitation. Similarly in electrodynamics the field strengths can be classified into two types: radiative and general.[2] Thus, it is desirable to have a reasonable classification of non-Abelian gauge fields to have a better insight of the nonlinear feature of non-Abelian gauge theory. Recently, the algebraic classification of classical non-Abelian gauge fields has been tried by several authors,[3-5] and an acceptable classification is proposed by Anandan and Tod.[6] Unfortunately, this classification is made basically according to their SL(2,C) configurations of the fields so that, as far as the dimension of the internal symmetry group G is no less than three, the classification does not distinguish the internal symmetry structure. On the other hand, since different groups have different algebraic structures, a proper classification should be able to reflect the algebraic structure of the internal symmetry group as well. The purpose of this paper is to present a general scheme to obtain such a classification. The idea we adopt here is to classify the fields according to the little groups of the symmetry groups SL(2,C) and G. The general procedure how to realize this idea is presented in the following and the groups SU(2) and SU(3) are considered in detail as explicit examples.

We will start from the SL(2,C) classification[6] and introduce the concept of the canonical form and the canonical frame for each different SL(2,C) class. As we will see soon, in the canonical forms all the SL(2,C) degrees of freedom are removed as far as possible so that basically only the internal degrees of freedom are left arbitrary in the canonical frame. Then with the help of these canonical forms we will answer the question of, given a group G, what are the possible little groups for each type of the canonical form, how many SL(2,C) $\otimes$ G inequivalent realizations of a given class are possible, and how one can realize them all in a systematic way. This leads us to a natural classification of the gauge fields which is free of the undesirable feature that the existing SL(2,C) classification has.

The paper is organized as follows: In Sec. II the SL(2,C) classification[6] is briefly reviewed for later convenience. In Sec. III we introduce the canonical form and the corresponding canonical frame for each of the five different SL(2,C) types of the gauge field strengths in spinor formalism. In Sec. IV the classification is refined according to the possible little groups of the field configurations under the internal symmetry group G. With the help of the canonical forms we also present a systematic way how to find out the number of all the physically inequivalent realizations of a given type, and how to construct them all. As explicit examples the groups SU(2) and SU(3) are examined in detail. Naturally, one obtains richer structure as the group G gets larger. However, already at the level of SU(2) the refinement is possible which enables us to separate the purely non-Abelian configurations from the SU(2) embeddings of the Abelian types. This way we will see that the simplest non-Abelian configuration is the dual type which has the local dual symmetry as a gauge symmetry. In the last section comments are made on various aspects, in particular on the local stability of the classification. We argue that the classification has a "minimal" local stability. Also, the generalization of the classification to the coupled Einstein–Yang–Mills system is briefly discussed.

II. SL(2,C) CLASSIFICATION: A BRIEF REVIEW

Non-Abelian gauge fields of any group G (of dimension no smaller than three may be classified into five major types according to their SL(2,C) configurations.[6] For the notational convenience we will briefly review the classification in this section.

First, as a preliminary remark remember that in SL(2,C) spinor formalism any real antisymmetric tensor $F_{\mu\nu}$ ($\mu,\nu = 0,1,2,3$) can be specified by a symmetric bispinor ϕ_{AB} ($A,B = 1,2$) and *vice versa*. The relation between $F_{\mu\nu}$, $F^*_{\mu\nu}$ and ϕ_{AB} can be given by

$$\begin{aligned} \phi_{AB} &= \tfrac{1}{2}\sigma^{\mu}_{\dot{c}A}\sigma_B^{\nu\dot{c}}F_{\mu\nu}, \\ F_{\mu\nu} &= 2\,\mathrm{Re}\sigma^{A}_{\mu\dot{c}}\sigma_{\nu}^{\dot{c}B}\phi_{AB}, \\ F^*_{\mu\nu} &= 2\,\mathrm{Im}\sigma^{A}_{\mu\dot{c}}\sigma_{\nu}^{\dot{c}B}\phi_{AB}, \end{aligned} \tag{1}$$

where

$$\sigma^{\mu}_{\dot{c}A} = \tfrac{1}{2}(1,\boldsymbol{\sigma})_{\dot{c}A}, \tag{2}$$

$\boldsymbol{\sigma}$ being the Pauli matrices.

Now, for any given group G of dimension n, the gauge field strengths $\mathbf{F}_{\mu\nu}$ in terms of the symmetric bispinors $\boldsymbol{\phi}_{AB}$ can be written as

$$\boldsymbol{\phi}_{AB} = \boldsymbol{\phi}\alpha_A\alpha_B + \boldsymbol{\chi}\beta_A\beta_B + \boldsymbol{\xi}\alpha_{(A}\beta_{B)}, \tag{3}$$

 0022-2488/79/122605-08$01.00

where

$$\phi = \phi_{AB}\,\beta^{A}\beta^{B},$$

$$\chi = \phi_{AB}\alpha^{A}\alpha^{B},$$

$$\xi = -2\phi_{AB}\alpha^{A}\beta^{B},$$

α_A and β_A are an SL(2,C) orthonormal spinor basis,

$$\alpha_A\,\beta^{A} = 1,$$

and

$$\alpha_{(A}\,\beta_{B)} = \tfrac{1}{2}(\alpha_A\,\beta_B + \alpha_B\,\beta_A).$$

Clearly, ϕ, χ, and ξ form adjoint representations of the symmetry group G. Furthermore, for each component i ($i = 1,2,\ldots,n$), ϕ^i,χ^i,ξ^i forms a (1,0) spinor representation of SL(2,C). In fact, one can easily derive the transformation laws of ϕ, χ. and ξ explicitly under a change of the SL(2,C) basis α_A and β_A. Under the SL(2,C) transformation

$$\begin{pmatrix}\alpha'_A\\ \beta'_A\end{pmatrix} = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}\alpha_A\\ \beta_A\end{pmatrix}, \quad ad - bc = 1, \tag{4}$$

the form invariance of the field strengths

$$\begin{aligned}\phi_{AB} &= \phi\alpha_A\alpha_B + \chi\beta_A\,\beta_B + \xi\alpha_{(A}\,\beta_{B)}\\ &= \phi'\alpha'_A\alpha'_B + \chi'\beta'_A\,\beta'_B + \xi'\alpha'_{(A}\,\beta'_{B)}\end{aligned} \tag{5}$$

requires us to have

$$\begin{pmatrix}\phi'\\ \chi'\\ \xi'\end{pmatrix} = \begin{pmatrix}d^2 & c^2 & -cd\\ b^2 & a^2 & -ab\\ -2bd & -2ac & ad+bc\end{pmatrix}\begin{pmatrix}\phi\\ \chi\\ \xi\end{pmatrix}. \tag{6}$$

Now consider the $3\times n$ matrix Φ:

$$\Phi = \begin{pmatrix}\phi', & \phi^2,\ldots,\phi^n\\ \chi', & \chi^2,\ldots,\chi^n\\ \xi', & \xi^2,\ldots,\xi^n\end{pmatrix} \tag{7}$$

and the $n\times n$ matrix L whose components L^{ij} are given by

$$\begin{aligned}L^{ij} &= \phi^{iB}_{A}\phi^{jA}_{B}\\ &= -\tfrac{1}{4}(F^{i}_{\mu\nu}F^{j\mu\nu} + iF^{*i}_{\mu\nu}F^{j\mu\nu})\\ &= \tfrac{1}{2}\xi^i\xi^j - (\phi^i\chi^j + \phi^j\chi^i).\end{aligned} \tag{8}$$

Clearly, the ranks of these matrices Φ and L are SL(2,C) $\otimes$ G invariant. Thus, one can classify the gauge fields ϕ_{AB} into the following five types according to the ranks of Φ and L:

(a) type I; rank $\Phi = 1$, rank $L = 0$;
(b) type II; rank $\Phi = 1$, rank $L = 1$;
(c) type III; rank $\Phi = 2$, rank $L = 1$;
(d) type IV; rank $\Phi = 2$, rank $L = 2$;
(e) type V; rank $\Phi = 3$, rank $L = 3$.

Here we have excluded the trivial case $\phi_{AB} = 0$. The fact that there exist no other types can easily be proved,[6] which will not be reconfirmed here.

III. CANONICAL FORMS

Now we will introduce the concept of the canonical forms and the canonical frames for the above five types of the fields. In the canonical frames all the SL(2,C) degrees of freedom of the corresponding field configurations are removed as far as possible so that only the internal degrees of freedom are left arbitrary. As we will see later the canonical forms and the corresponding canonical frames turn out to be extremely useful when we consider the SL(2,C) $\otimes$ G inequivalent realizations of a given class of the fields.

Let us show that each of the five types described above can be put into the corresponding canonical form which we specify in the following:

(a) type V; rank $\Phi = 3$, rank $L = 3$: In this case the canonical form can be chosen as

$$\phi_{AB} = \phi\alpha_A\alpha_B + \chi\beta_A\,\beta_B + \xi\alpha_{(A}\,\beta_{B)}, \tag{9a}$$

$$\alpha_A\,\beta^{A} = 1,$$

where ϕ, χ, and ξ are linearly independent (under complex coefficients). Our choice of the form (9a) is obvious. The fact that rank $\Phi = 3$ guarantees that rank $L = 3$. Notice, however, that as it stands the form (9a) does not specify the SL(2,C) basis so that the full six degrees of freedom are still left arbitrary there. One can remove this arbitrariness by imposing *any* six mutually independent and G-invariant constraints on ϕ, χ, and ξ as far as possible. However, if the multiplets satisfy[7]

$$\phi^2 = 0, \quad \chi = \phi^*$$

$$\phi\times\chi \neq 0, \quad \xi\times(\phi\times\chi) = 0, \tag{9b}$$

one may not always be able to fix all the SL(2,C) degrees of freedom by G-invariant constraints. This can actually happen when the form (9a) admits a little group of the mixed type, namely, a little group consisting of a simultaneous SL(2,C) and the internal group transformations.[8] We will analyze this possibility case by case when we consider the internal structure in the following section. With this exception all the SL(2,C) degrees of freedom can be fixed completely in the canonical frame.

(b) Type IV; rank $\Phi = 2$, rank $L = 2$: In this case one can always choose

$$\phi_{AB} = \phi\alpha_A\alpha_B + \chi\beta_A\,\beta_B, \quad \alpha_A\,\beta^{A} = 1. \tag{10a}$$

Here, of course, ϕ and χ are linearly independent. The SL(2,C) degrees of freedom can be fixed completely by G-invariant constraints except when

$$\phi^2 = 0, \quad \chi \simeq \phi^*, \quad \phi\times\chi \neq 0. \tag{10b}$$

For example, if $\phi^2 \neq 0$ (or $\chi^2 \neq 0$), the SL(2,C) degrees of freedom can be fixed completely by requiring

$$\phi^2 = 1 \quad (\text{or } \chi^2 = 1). \tag{10c}$$

If the condition (10b) is satisfied, however, the SL(2,C) degrees of freedom may not always be fixed completely in the canonical frame, and the form (10a) may admit a little group of the mixed type as we will see in the following section.

Let us prove the statement. Clearly, if ϕ_{AB} is given by Eq. (10a), one has rank Φ = rank $L = 2$. To show that the inverse is also true, or to show that one can always choose the form (10a) whenever rank Φ = rank $L = 2$, let us write ϕ_{AB} in its most general form (3):

$$\phi_{AB} = \phi\alpha_A\alpha_B + \chi\beta_A\,\beta_B + \xi\alpha_{(A}\,\beta_{B)}, \quad \alpha_A\,\beta^{A} = 1.$$

Now, since rank $\Phi = 2$ by assumption, one must have

$$x\phi + y\chi + z\xi = 0 \tag{11}$$

for some complex x, y, and z, not all of which vanish. Obvi-

ously, there are two possibilities, either $z\neq 0$ or else $z=0$. Consider the case $z\neq 0$ first. Then

$$\xi = -\frac{x}{z}\phi - \frac{y}{z}\chi.$$

This implies that ϕ and χ must be linearly independent. Then notice that the $n\times n$ matrix L is given by

$$L^{ij} = \frac{1}{2}\frac{x^2}{z^2}\phi^i\phi^j + \frac{1}{2}\frac{y^2}{z^2}\chi^i\chi^j - \left(1-\frac{1}{2}\frac{xy}{z^2}\right)(\phi^i\chi^j + \phi^j\chi^i).$$

Writing

$$\phi = \begin{pmatrix}\vdots\\ p\\ \vdots\\ q\\ \vdots\end{pmatrix}, \chi = \begin{pmatrix}\vdots\\ s\\ \vdots\\ t\\ \vdots\end{pmatrix},$$

one can easily show that the determinant of the 2×2 submatrix L_2 made of p,q and s,t components is given by

$$\det L_2 = -(1-xy/z^2)(pt-qs)^2.$$

Since one may assume (remember that ϕ and χ are linearly independent by assumption) that

$$pt - qs\neq 0,$$

the condition that rank $L=2$ requires us to have

$$z^2 - xy\neq 0 \tag{12a}$$

when $z\neq 0$.

Next, consider the other case $z=0$. Then the condition (11) reduces to

$$x\phi + y\chi = 0,$$

where not both of x and y vanish. Without loss of generality one may further assume that $y\neq 0$, or

$$\chi = -\frac{x}{y}\phi$$

in which case one has

$$L^{ij} = \tfrac{1}{2}\xi^i\xi^j - 2\frac{x}{y}\phi^i\phi^j.$$

Then, with

$$\phi = \begin{pmatrix}\vdots\\ p\\ \vdots\\ q\\ \vdots\end{pmatrix} \quad \xi = \begin{pmatrix}\vdots\\ s\\ \vdots\\ t\\ \vdots\end{pmatrix},$$

one can show that the determinant of the 2×2 submatrix L_2 made of p,q and s,t components is

$$\det L_2 = -\frac{x}{y}(pt-qs)^2.$$

Thus, when $z=0$, the condition that rank $L=2$ requires us to have

$$xy\neq 0. \tag{12b}$$

In short, one has rank $L=2$ if and only if

$$z^2 - xy\neq 0$$

or

$$z^2 - xy = 1 \tag{12'}$$

by choosing a proper normalization. Then it is easy to see that one can always find an SL(2,C) transformation (6) in which one has

$$\xi' = -2bd\phi - 2ac\chi + (ad+bc)\xi = 0. \tag{13}$$

This is so since given x, y, and z there always exist a, b, c, and d which satisfy

$$x = -2bd,$$
$$y = -2ac,$$
$$z = ad + bc,$$

with

$$z^2 - xy = (ad-bc)^2 = 1.$$

This completes the proof that whenever rank Φ = rank $L = 2$ one can always choose the form (10a).

Notice, however, that the condition (13) determines only four prameters of SL(2,C) so that there still remain two parameter degrees of freedom in (10a). This can easily be seen by noticing that the form invariance of Eq. (10a), i.e.,

$$\phi_{AB} = \phi\alpha_A\alpha_B + \chi\beta_A\beta_B = \phi'\alpha'_A\alpha'_B + \chi'\beta'_A\beta'_B, \tag{14}$$

allows the following SL(2,C) degrees of freedom:

$$\text{(i)}\quad \begin{pmatrix}\alpha'_A\\ \beta'_A\end{pmatrix} = \begin{pmatrix}a & 0\\ 0 & \frac{1}{a}\end{pmatrix}\begin{pmatrix}\alpha_A\\ \beta_A\end{pmatrix}, \quad \begin{pmatrix}\phi'\\ \chi'\end{pmatrix} = \begin{pmatrix}\frac{1}{a^2} & 0\\ 0 & a^2\end{pmatrix}\begin{pmatrix}\phi\\ \chi\end{pmatrix} \tag{15a}$$

or

$$\text{(ii)}\quad \begin{pmatrix}\alpha'_A\\ \beta'_A\end{pmatrix} = \begin{pmatrix}0 & \frac{1}{c}\\ -c & 0\end{pmatrix}\begin{pmatrix}\alpha_A\\ \beta_A\end{pmatrix}, \quad \begin{pmatrix}\phi'\\ \chi'\end{pmatrix} = \begin{pmatrix}0 & c^2\\ \frac{1}{c^2} & 0\end{pmatrix}\begin{pmatrix}\phi\\ \chi\end{pmatrix}, \tag{15b}$$

which together form a two parameter subgroup of SL(2,C). Clearly, one can remove these degrees of freedom by imposing the constraint (10c) or similar ones, except when Eq. (10b) is satisfied. With this exception all the SL(2,C) degrees of freedom can be fixed in the canonical frame.

(c) Type III; rank $\Phi = 2$, rank $L = 1$: In this case one can always choose the following canonical form:

$$\phi_{AB} = \phi\alpha_A\alpha_B + \xi\alpha_{(A}\beta_{B)}, \quad \alpha_A\beta^A = 1. \tag{16a}$$

Here again ϕ and ξ are linearly independent. The SL(2,C) degrees of freedom can be fixed completely in the canonical frame by imposing G-invariant constraints on ϕ and ξ except when

$$\phi^2 = 0, \quad \phi^*\times\phi\neq 0, \quad \xi\times(\phi^*\times\phi) = 0. \tag{16b}$$

TABLE I. The SL(2,C) classification of non-Abelian gauge fields, and their canonical forms. The ϕ, χ, and ξ are linearly independent and the spinor basis is orthonormal. The local stability of the classification is indicated by the arrows, against which it is stable.

Rank L \ Rank Φ	0	1	2	3	SL(2,C) type	canonical from ϕ_{AB}
0	0 ← I				I	$\phi\alpha_A\alpha_B$
1		II	III		II	$\xi\alpha_{(A}\beta_{B)}$
					III	$\phi\alpha_A\alpha_B + \xi\alpha_{(A}\beta_{B)}$
2			IV		IV	$\phi\alpha_A\alpha_B + \chi\beta_A\beta_B$
3				V	V	$\phi\alpha_A\alpha_B + \chi\beta_A\beta_B + \xi\alpha_{(A}\beta_{B)}$

For example, when $\phi^2\xi^2 \neq (\phi\cdot\xi)^2$, one may require

$$\phi^2 = 1, \quad \phi\cdot\xi = 0 \quad (\text{if } \xi^2 \neq 0), \tag{16c}$$
$$\phi^2 = 0, \quad \phi\cdot\xi = 1 \quad (\text{if } \xi^2 = 0)$$

to fix the SL(2,C) degrees of freedom completely.

The proof that one can always choose the form (16a) is simple. From the preceding argument to obtain the canonical form (10) one finds that the condition that rank $\Phi = 2$ and rank $L = 1$ requires

$$z^2 - xy = 0. \tag{17}$$

Then one can easily convince oneself that there always exists an SL(2,C) transformation (6) in which one has

$$\chi' = b^2\phi + a^2\chi - ab\xi = 0. \tag{18}$$

Hence, one is led to the form (16a). Notice, however, that the condition (18) removes only two SL(2,C) degrees of freedom. The remaining four degrees of freedom can be fixed by imposing G invariant constraints like Eq. (16c) except when Eq. (16b) is satisfied. Again, barring the exceptional case, all the SL(2,C) degrees of freedom can be fixed in the canonical frame.

(d) Type II; rank $\Phi = 1$, rank $L = 1$: In this case the canonical form is given by

$$\phi_{AB} = \xi\alpha_{(A}\beta_{B)}, \quad \alpha_A\beta^A = 1, \tag{19}$$

with no constraint on ξ. The fact that the above canonical form exists comes from the fact that when rank $\Phi = 1$ the system reduces essentially (although not exactly) to Abelian configurations as far as the SL(2,C) structure is concerned. As in the Abelian case the canonical form (19) fixes four SL(2,C) degrees of freedom. The remaining two degrees of freedom cannot be fixed by any G-invariant constraint on ξ since the form invariance of (19) allows the following SL(2,C) degrees of freedom:

$$\text{(i)} \quad \begin{pmatrix} \alpha'_A \\ \beta'_A \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix} \begin{pmatrix} \alpha_A \\ \beta_A \end{pmatrix}, \quad \xi' = \xi \tag{20a}$$

or

$$\text{(ii)} \quad \begin{pmatrix} \alpha'_A \\ \beta'_A \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{c} \\ -c & 0 \end{pmatrix} \begin{pmatrix} \alpha_A \\ \beta_A \end{pmatrix}, \quad \xi' = -\xi, \tag{20b}$$

which together form a two parameter little group B_2 of SL(2,C) of the canonical form (19).

(e) Type I; rank $\Phi = 1$, rank $L = 0$: Clearly, one can choose

$$\phi_{AB} = \phi\alpha_A\alpha_B. \tag{21a}$$

Again, as in the Abelian case, the form (21a) fixes two SL(2,C) degrees of freedom. One can remove two more by imposing G-invariant constraints on ϕ except the case

$$\phi^2 = 0, \quad \phi^* \times \phi \neq 0. \tag{21b}$$

For example, if $\phi^2 \neq 0$ the constraint

$$\phi^2 = 1 \tag{21c}$$

can fix two more parameters of SL(2,C) in the canonical frame. However, there always remains the following two parameter SL(2,C) little group A_2 of Eq. (21):

$$\begin{pmatrix} \alpha'_A \\ \beta'_A \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} \alpha_A \\ \beta_A \end{pmatrix}. \tag{22}$$

When the condition (21b) is satisfied, the form (21a) may admit a little group of the mixed type.

This completes our introduction of the canonical forms and the canonical frames. The results are summarized in Table I.

IV. INTERNAL STRUCTURE

In the above we have examined in detail the SL(2,C) structure of the fields. Now we turn to the internal symmetry structure and ask that, given a group G, how many algebraically different G structures are possible for each of the above five types, how many SL(2,C) $\otimes$ G inequivalent realizations are allowed, and how one can construct them all in a systematic way. To answer the first question, however, one has to first decide what one means by the algebraically different G structures since, as will become clear, there are in general infinitely many SL(2,C) $\otimes$ G inequivalent realizations of the fields for each of the above five types. We find it natural to distinguish two field configurations when they have different little groups of G. So we will call two configurations algebraically different when they admit different little groups of G. This convention has a merit of providing us a natural way to figure out the unbroken symmetry structure of a given configuration or the way how a symmetry is embedded into a larger one.

With this convention we can now answer the questions raised above completely. The general way to proceed to answer the questions is clear. First, remove all the SL(2,C) degrees of freedom as far as possible by choosing the canoni-

cal frame. Then find out all the possible little groups of G allowed by the configuration of ϕ, χ, and ξ. For this the fact that in the canonical frame ϕ, χ, and ξ can be considered as SL(2,C) scalars barring the few exceptional cases will be extremely helpful. We will now apply this procedure to the groups SU(2) and SU(3) as explicit examples.

A. SU(2)

Obviously only one nontrivial little group H is possible, which is U(1). Let us consider the five different types separately.

1. Type I

Remember that this type admits a two parameter little group A_2 [i.e., Eq. (22)] of SL(2,C). Writing

$$\phi = \phi_1 + i\phi_2 \quad (\phi_1, \phi_2 \text{ real}), \tag{23}$$

one can always find a canonical frame in which

$$\phi_1^2 - \phi_2^2 = 1 \quad (\text{or } 0), \qquad \phi_1 \cdot \phi_2 = 0. \tag{21b'}$$

Notice that the real and the imaginary components of ϕ can always be made orthogonal in the canonical frame. Now, there are the following two possibilities:

(a) I_1; $H = U(1)$: In the canonical frame this is possible only if ϕ is real (i.e., $\phi_2 = 0$). Furthermore, due to the SU(2) degrees of freedom one can always choose

$$\phi = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{24}$$

Thus, there exist only one SL(2,C) $\otimes$ G inequivalent realization of this type. Obviously, this is the Abelian subclass embedded in SU(2).

(b) I_2; H is trivial: In this case one can always put

$$\phi = \begin{pmatrix} \phi_1 \\ -i\phi_2 \\ 0 \end{pmatrix}, \quad \phi_1^2 - \phi_2^2 = 1 \ (\text{or } 0) \tag{25}$$

so that

$$\mathbf{F}_{\mu\nu} = \begin{pmatrix} F_{\mu\nu} \\ \dfrac{\phi_2}{\phi_1} F^*_{\mu\nu} \\ 0 \end{pmatrix}, \tag{25'}$$

where

$$F_{\mu\nu} = 2\phi_1 \, \mathrm{Re}\sigma^A_{\mu\dot{c}C}\sigma_\nu^{\dot{c}B}\alpha_A\alpha_B.$$

Clearly, all possible SL(2,C) $\otimes$ G inequivalent realizations of this type form a one-dimensional manifold R^1.

This class deserves to be noted in two respects. First, this is the simplest possible non-Abelian configuration. Second, the two nonvanishing components of $\mathbf{F}_{\mu\nu}$ are dual to each other up to a scalar factor so that when $\phi^2 = 0$ the configuration has the following local dual symmetry:

$$\begin{pmatrix} F'_{\mu\nu} \\ F^{*\prime}_{\mu\nu} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} F_{\mu\nu} \\ F^*_{\mu\nu} \end{pmatrix}, \tag{26}$$

not only as an SU(2) but also as an SL(2,C) transformation. This means that this configuration admits a little group U(1) of the mixed type. Notice that this is indeed the exceptional case (21b). Since three parameters of SL(2,C) (including the two parameters of the little group A_2) cannot be fixed in the canonical frame there exists only one SL(2,C) $\otimes$ G inequivalent realization of this exceptional case.

2. Type II

This type also allows a two parameter little group B_2 [i.e., Eq. (20)] of SL(2,C). For the internal symmetry one still finds two possibilities:

(a) II_1; $H = U(1)$: Since we have no constraint on ξ, the possible SL(2,C) $\otimes$ G inequivalent realizations form a two-dimensional manifold R^2. Clearly, this is the embedding of the Abelian nonradiative configurations into SU(2).

(b) II_2; H is trivial: In this case one can use the full SU(2) degrees of freedom to remove three out of six parameters of ξ, so that the SL(2,C) $\otimes$ G inequivalent realizations form an R^3.

3. Type III

Remember that this type breaks SL(2,C) completely. Furthermore, due to the linear independence of ϕ and ξ, there can be no nontrivial little group of SU(2). In short, both SL(2,C) and G are completely broken in this case. Among the 12 parameters of ϕ and ξ, seven can be fixed by the three SU(2) degrees of freedom and four SL(2,C) constraints like Eq. (16c). Thus, the SL(2,C) $\otimes$ G inequivalent realizations form an R^5. However, if Eq. (16b) is satisfied, there remains one parameter of SL(2,C) which cannot be fixed by a G-inequivalent constraint in the canonical frame. In this case one can always put

$$\phi = \hat{e}_1 + i\hat{e}_2, \quad \xi = \rho e^{i\alpha}\hat{e}_1 \times \hat{e}_2, \qquad \hat{e}_1^2 = \hat{e}_2^2 = 1, \quad \hat{e}_1 \cdot \hat{e}_2 = 0, \tag{27}$$

by a proper choice of an SL(2,C) basis. Clearly, the canonical form admits a little group U(1) of the mixed type, and the SL(2,C) $\otimes$ G inequivalent realizations are reduced to an R^2.

4. Type IV

Again both SL(2,C) and G are totally broken. Two SL(2,C) constraints like Eq. (10c) and three SU(2) degrees of freedom can be used to remove five parameters of ϕ and χ, so that the SL(2,C) $\otimes$ G inequivalent realizations form an R^7. When ϕ and χ satisfy Eq. (10b), however, they can be brought into the form

$$\phi = \frac{\rho}{\sqrt{2}} e^{i(\alpha/2)}(\hat{e}_1 + i\hat{e}_2), \qquad \chi = \frac{\rho}{\sqrt{2}} e^{i(\alpha/2)}(\hat{e}_1 - i\hat{e}_2) \tag{28}$$

by a proper choice of an SL(2,C) basis. Here again the SL(2,C) degrees of freedom cannot be fixed completely in the canonical frame since Eq. (28) admits a little group U(1) of the mixed type. The SL(2,C) $\otimes$ G inequivalent realizations are then reduced to an R^2.

TABLE II. The SL(2,C) ⊗ SU(2) classification.

Type	Little group SL(2,C)	Little group SU(2)	Number of algebraically inequivalent realizations	Remarks
I_1	A_2	U(1)	1	Abelian
I_2	A_2	I	R^1	SU(2) (dual)
II_1	B_2	U(1)	R^2	Abelian
II_2	B_2	I	R^3	SU(2)
III	I	I	R^5	SU(2)
IV	I	I	R^7	SU(2)
V	I	I	R^9	SU(2)

5. Type V

Obviously, the full nine SL(2,C) ⊗ G degrees of freedom can be used to remove the superficial degrees in ϕ, χ, and ξ, so that the physically inequivalent realizations form an R^9. If the multiplets ϕ, χ, and ξ satisfy Eq. (9b), however, not all the SL(2,C) degrees of freedom can be fixed in the canonical frame. In fact, if Eq. (9b) is satisfied, one may put

$$\phi = \frac{\rho}{\sqrt{2}} e^{i(\alpha/2)}(\hat{e}_1 + i\hat{e}_2),$$
$$\chi = \frac{\rho}{\sqrt{2}} e^{i(\alpha/2)}(\hat{e}_1 - i\hat{e}_2), \tag{29}$$
$$\xi = \rho' e^{i(\alpha'/2)}\hat{e}_1 \times \hat{e}_2,$$

so that the canonical form admits a little group U(1) of the mixed type. Thus, the SL(2,C) ⊗ G inequivalent realizations are reduced to an R^4. In a more special case when

$$\xi = i\frac{\phi \times \chi}{\sqrt{\phi \cdot \chi}} \tag{30}$$

is satisfied, one has $\rho' = \rho$ and $\alpha' = \alpha$ so that the canonical form admits a little group SU(2) of the mixed type [i.e., a mixture of two SU(2) subgroups of SL(2,C) and SU(2)]. This means that in this case three parameters of SL(2,C) are left arbitrary in the canonical frame. Then the SL(2,C) ⊗ G inequivalent realizations are further reduced to an R^2.

This completes the classification for SU(2). The results are summarized in Table II.

B. SU(3)

In this case three nontrivial little groups of G, i.e., SU(2) ⊗ U(1), U(1) ⊗ U(1), and U(1), are possible. Naturally one expects a richer structure here. Since the SL(2,C) structure will remain the same as before, we will concentrate on the G structure.

1. Type I

(a) I_1; H = SU(2) ⊗ U(1): Obviously this is an Abelian subclass. Again, due to the constraint (21c), any realizations of this type are equivalent to each other. Notice, however, that this is *not* the SU(3) embedding of the SU(2) Abelian subclass, which must necessarily have U(1) ⊗ U(1) as its little group. The reason why one can get different Abelian embeddings in this case is, of course, due to the fact that SU(3) has rank two.

(b) I_2; H = U(1) ⊗ U(1): In this case one can always put

$$\phi^i = \delta^i_3(\alpha + i\beta) + \delta^i_8(\gamma - i\delta) \quad (\alpha,\beta,\gamma,\delta \text{ real}), \tag{31}$$

with

$$\alpha\beta = \gamma\delta,$$
$$\alpha^2 + \gamma^2 - (\beta^2 + \delta^2) = 1 \text{ (or } 0),$$

so that the SL(2,C) ⊗ G inequivalent realizations form an R^2. Notice that this class includes the SU(3) embedding of the SU(2) Abelian subclass (with $\beta = \gamma = \delta = 0$), or the SU(3) embedding of two Abelian fields (with $\beta = \delta = 0$).

(c) I_3; H = U(1): In this case only five parameters of ϕ can be considered G inequivalent, which can further be reduced to three by two SL(2,C) constraints. Thus, the physically inequivalent realization of this class forms an R^3. Notice that the SU(3) embedding of the corresponding SU(2) type I_2 belongs to this class. Thus, a little group U(1) of the mixed type is possible as the SU(3) embedding of the corresponding SU(2) type. However, observe that the condition (21b) does not always guarantee the existence of the little groups.

(d) I_4; H is trivial: The full SU(3) degrees of freedom and the constraint (21c) can be used to remove the unphysical degrees of freedom. Thus, the SL(2,C) ⊗ G inequivalent realizations form an R^6. Notice that the SU(2) type I_2 can also appear here as a different embedding, namely, the "embedding" of SU(2) into the maximal subgroup SO(3) of SU(3). This is possible since the field strengths form an adjoint representation.

2. Type II

Again there exist four different classes II_1, II_2, II_3, and II_4 with the little groups SU(2) ⊗ U(1), U(1) ⊗ U(1), U(1), and trivial, respectively. The SL(2,C) ⊗ G inequivalent realizations form the manifold of R^2, R^4, R^5, and R^8, respectively. Notice that one cannot obtain the class II_1 from any SU(3) embedding of the corresponding SU(2) types.

3. Type III

In this case the little group SU(2) ⊗ U(1) is forbidden due to the linear independence of ϕ and ξ, so that there are

three subclasses III_1, III_2, and III_3 with the little group $U(1) \otimes U(1)$, $U(1)$, and trivial, respectively. The physically inequivalent realizations form R^4, R^9, R^{20}, respectively. Again, III_1 represents a genuine SU(3) subclass. Also, both III_2 and III_3 allow the embedding of the SU(2) type III. If Eq. (16b) is satisfied, however, one may put

$$\phi = \hat{e}_1 + i\hat{e}_2,$$
$$\xi = \rho e^{i\alpha}\hat{e}_1 \times \hat{e}_2 + \rho' e^{i\alpha'}(\hat{e}_1 \times \hat{e}_2) * (\hat{e}_1 \times \hat{e}_2). \quad (32)$$

where $*$ is meant to be the symmetric product (the d product) of SU(3). Consequently, the canonical form admits a little group U(1) of the mixed type, and the physically inequivalent realizations reduce to an R^4. This is possible in both III_2 and III_3.

4. Type IV

Again the little group $SU(2) \otimes U(1)$ is forbidden. The three subclasses IV_1, IV_2, and IV_3 with the little group $U(1) \otimes U(1)$, $U(1)$, and trivial form R^6, R^{11}, and R^{22}, respectively. Again IV_1 forms a genuine SU(3) subclass. Both IV_2 and IV_3 may admit a little group U(1) of the mixed type as the SU(3) embedding of the corresponding SU(2) type.

5. Type V

In this case two of the little groups $SU(2) \otimes U(1)$ and $U(1) \otimes U(1)$ are forbidden. The two subclasses V_1 and V_2 with the little groups U(1) and trivial form the $SL(2,C) \otimes G$ inequivalent manifold of R^{15} and R^{34}, respectively. When Eq. (9b) is satisfied one may put

$$\phi = \frac{\rho}{\sqrt{2}} e^{i(\alpha/2)}(\hat{e}_1 + i\hat{e}_2),$$
$$\chi = \frac{\rho}{\sqrt{2}} e^{i(\alpha/2)}(\hat{e}_1 - i\hat{e}_2), \quad (33)$$
$$\xi = \rho' e^{i(\alpha'/2)}\hat{e}_1 \times \hat{e}_2 + \rho'' e^{i(\alpha''/2)}(\hat{e}_1 \times \hat{e}_2) * (\hat{e}_1 \times \hat{e}_2).$$

This configuration admits a little group U(1) of the mixed type. Then the physically inequivalent realizations are reduced to an R^6. This is possible in both V_1 and V_2. If one further has Eq. (30), the field configuration admits a little group SU(2) of the mixed type, and the physically inequivalent realizations are further reduced to an R^2. Again this is possible in both V_1 and V_2.

This completes our analysis for SU(3). The results are summarized in Table III.

V. DISCUSSIONS

We have shown a general way how, given an internal symmetry group G, one can systematically classify the field configurations according to their algebraic structure, and how one can find all the $SL(2,C) \otimes G$ inequivalent realizations of a given class. Although explicit demonstration is carried out for the groups SU(2) and SU(3), it is clear that the procedure can be applied systematically to any group. To carry out the procedure the concept of the canonical form and the canonical frame is extremely useful.

In principle, one can choose a different classification.[8] However, ours has the merit of showing us most clearly the unbroken symmetry structure of a given type of the fields, or the embedding structure of a symmetry into a larger one.

Now we should like to make a few comments on the classification. First, the classification has been made according to the algebraic properties of the field strengths $F_{\mu\nu}$ (or ϕ_{AB}), and not in terms of the gauge potentials A_μ. Thus, the differential properties of the potentials (as well as the fields) have *not* been taken into account. So, for example, the question whether there always exist gauge potentials A_μ that could yield the field strengths of any given class considered above in a *finite* region of space–time has yet to be answered. Here the situation is similar to the Petrov classification of the gravitational fields, where the classification is made in terms of the curvature tensor, not in terms of the metric tensor.

Next, the above classification is local in the sense that it applies to each space–time point. Now if a field configuration belongs to a given class at a point, one might wonder whether this guarantees that the configuration belongs to the

TABLE III. The $SL(2,C) \otimes SU(3)$ classification.

Type	Little group SL(2,C)	Little group SU(3)	Number of algebraically inequivalent realizations	Remarks
I_1	A_2	$SU(2) \otimes U(1)$	1	Abelian [SU(3)]
I_2	A_2	$U(1) \otimes U(1)$	R^2	Embedding, SU(3)
I_3	A_2	U(1)	R^3	Embedding, SU(3)
I_4	A_2	I	R^6	Embedding, SU(3)
II_1	B_2	$SU(2) \otimes U(1)$	R^2	Abelian [SU(3)]
II_2	B_2	$U(1) \otimes U(1)$	R^4	Embedding, SU(3)
II_3	B_2	U(1)	R^5	Embedding, SU(3)
II_4	B_2	I	R^8	Embedding, SU(3)
III_1	I	$U(1) \otimes U(1)$	R^4	SU(3)
III_2	I	U(1)	R^9	Embedding, SU(3)
III_3	I	I	R^{20}	Embedding, SU(3)
IV_1	I	$U(1) \otimes U(1)$	R^6	SU(3)
IV_2	I	U(1)	R^{11}	Embedding, SU(3)
IV_3	I	I	R^{22}	Embedding, SU(3)
V_1	I	U(1)	R^{15}	Embedding, SU(3)
V_2	I	I	R^{34}	Embedding, SU(3)

same class in a certain neighborhood of the point (i.e., the question of the local stability of the classification). Unfortunately, this will not be the case in general. However, the classification *does* have a local stability in a limited sense. This can be explained in the following: If one makes an *arbitrary* yet infinitesimally small perturbation of the fields in a neighborhood of a point at which they belong to a given class, it becomes intuitively clear that one might be able to change the class to more complicated ones, but *not* to simpler ones at neighboring points. So, for example, a field configuration which belongs to the class IV at the origin could belong to the class V at some points contained in an arbitrarily small neighborhood of the origin, but there *always* exists a small neighborhood in which the field configuration does not belong to the class II or III. As a consequence the class V enjoys the complete local stability. Thus, one arrives at a one-way partial (or "minimal") local stability for the classification. In Table I the stability is indicated by the arrows against which the classification is stable. We leave the issue with the remark that the situation on this also remains the same as in Petrov classification.

Finally, it has been known that gauge theory itself can be viewed as Einstein's theory of gravitation in a higher dimensional space[9] in which the external four-dimensional space–time is unified with the n-dimensional internal space. In this unified geometry the internal symmetry group plays the role of the isometry group of the unified Riemannian space.[10] The resulting Einstein's theory in this higher dimensional space becomes the coupled Einstein–Yang–Mills theory after projected out to the four-dimensional space–time. Thus, one might wonder what kind of classification one could obtain for this unified theory. In this case one will have simply the direct product of the two classifications, i.e., the above classification for the gauge fields and the Petrov Classification for the gravitational fields. The reason that one does not get anything new is clear: First, the presence of the gauge fields restricts only the Ricci tensor, but not the Weyl tensor which is relevant for the Petrov classification. Then, if one assumes the Killing metric for the internal space, the Ricci tensor becomes traceless and algebraically trivial. Secondly, the introduction of the space–time curvature (i.e., the gravitation) does not affect the local algebraic structure of the gauge fields. The generalization of the classification further to the case in which the internal metric is no more of the Killing form but becomes scalar fields[9,10] can be made straightforward.

ACKNOWLEDGMENTS

It is a great pleasure to thank Dr. D. Maison and Dr. S. Schlieder for illuminating discussions, in particular on the local stability of the classification, and S.O. Ahn for the encouragement. I am also indebted to Dr. P. Breitenlohner for discussions.

[1]A.Z. Petrov, Sci. Nat. Kazan State Univ. **114**, 55 (1954).
[2]See, for example, J.L. Synge, *Relativity* (North-Holland, Amsterdam, 1956).
[3]T. Eguchi, Phys. Rev. D **13**, 1561 (1976); R. Roskies, Phys. Rev. D **15**, 1722 (1977).
[4]M. Carmeli, Phys. Rev. Lett. **39**, 523 (1977); J. Anandan and R. Roskies, Phys. Rev. D **18**, 1152 (1978).
[5]L.L. Wang and C.N. Yang, Phys. Rev. D **17**, 1687 (1978).
[6]J. Anandan and K.P. Tod, Phys. Rev. D **18**, 1144 (1978); J. Anandan, J. Math. Phys. **20**, 260 (1979).
[7]In this paper we use the Cartan–Killing metric for the internal space, assuming that the group G is semisimple.
[8]Recently, another classification of the SU(2) gauge field has been proposed by L. Castillejo, M. Kugler, and R. Roskies, where little groups of the mixed type have been considered seriously; see L. Castillejo, M. Kugler, and R. Roskies, Phys. Rev. D **19**, 1782 (1979).
[9]Y.M. Cho, J. Math. Phys. **16**, 2029 (1975); Y.M. Cho and P.G.O. Freund, Phys. Rev. D **12**, 1711 (1975).
[10]Y.M. Cho and P.S. Jang, Phys. Rev. D **12**, 3789 (1975).

Classical non-Abelian gauge theories in the space of reference frames

G. Cognola, R. Soldati, L. Vanzo, and S. Zerbini
Dipartimento di Matematica e Fisica della Libera Università di Trento, 38050—Povo (Trento), Italy

(Received 8 March 1979; accepted for publication 18 May 1979)

Classical non-Abelian gauge theories in presence of gravitation are reformulated in the language of the space of local reference frames. The geometric structure of this space, the transformation properties of the fields, and the possibility of introducing a covariant derivative are not assumed *a priori*, as in the usual classical theory, but they are deduced from a Lagrangian form and an action principle. Some examples of Lagrangian forms and the related equations of motion for the gauge fields and the matter fields are discussed.

1. INTRODUCTION

The idea that the space of reference frames plays a more fundamental role than the space–time in field theory was first considered by Lurçat.[1] This point of view arises in a natural way, if one thinks that the components of a tensor field are determined when their values at the space–time points are assigned with respect to a given local reference frame. Furthermore, if the fields are functions defined only on the space–time points, the concept of mass becomes more important than the concept of spin; on the other hand, it seems that there is not a deep reason, on a fundamental level, to believe in such a difference between them. These considerations led Lurçat to build a quantum field theory on the Poincaré group.

In a recent paper, Toller[2] developed a general formalism for classical field theory, where the fields were defined on a ten-dimensional manifold $\mathscr{S}$, namely the space of all the local reference frames; in this work he also succeeded in giving some examples of Lagrangian theories describing gravitation and electromagnetism. Afterwards Toller and Vanzo[3] considered the classical theory of free fields in this new language.

Recently a theory of gravity and supergravity on the Poincaré group has been developed by Ne'eman and Regge,[4] with a different approach. Our aim is to reformulate in the formalism of the space of reference frames a classical non-Abelian gauge theory in the presence of matter fields, taking into account gravitation. This may be achieved by generalizing the definition of local reference frame; in this way the dimension of the manifold $\mathscr{S}$ will become equal to $10 + N$, if N is the dimension of the gauge group, and the gauge fields will be described by N vector fields on $\mathscr{S}$, corresponding to infinitesimal gauge transformations of the first kind in the space $\mathscr{S}$. On the other side, gravitation will be described by ten-vector fields corresponding to the other infinitesimal transformations of the generalized reference frame. Thus, gravitation and gauge fields are unified at a geometrical level.

We recall that this basic idea is not new, but was first considered by Kaluza and Klein[5] in their five-dimensional unified theories of gravitation and electromagnetism, and was further extended to the non-Abelian case by several authors.[6]

The main difference between these approaches and ours is the following: In the formalism of the space of reference frames, the structure of the geometry of $\mathscr{S}$, the transformation properties of the geometric fields and the possibility to introduce covariant derivatives are deduced from an action principle and a Lagrangian form, while in the above mentioned treatments they have to be assumed *a priori*.

In Sec. 2 we give a brief review of the general formalism, Sec. 3 is devoted to the discussion of the Lagrangian form describing gravitation and Yang–Mills theories, and finally, in Sec. 4, we consider some examples of Lagrangian forms for matter fields, and we discuss some aspects of the conservation laws.

2. GENERAL FORMALISM

In this section we summarize the results given by Toller in Ref. 2. The fundamental framework where the physical observables are defined is no longer the space–time manifold $\mathscr{M}$, as in the usual theories, but the space $\mathscr{S}$ of all the local reference frames.

We say that a local reference frame is given if we fix the origin, the directions of the axis and a choice of N "gauge" parameters at the origin; thus we identify $\mathscr{S}$ with a n-dimensional differentiable manifold, where $n = 10 + N$. A mapping of $\mathscr{S}$ into itself will be called a transformation if it determines an equivalence class of transformation procedures, in the operational language,[2] that permit the building of a new reference frame starting from a preexisting one.

The infinitesimal transformations will be vector fields on $\mathscr{S}$ and we assume that they form an n-dimensional vector space $\mathscr{T}$. Let us denote by $\{A_\alpha\} = \{A_\alpha(s)\}$, $s \in \mathscr{S}$, $\alpha = 0,\ldots,n-1$ a basis in $\mathscr{T}$; we shall indicate by L_α the Lie derivative with respect to the vector field A_α, by i_α the corresponding inner product operator, and by $\{\omega^\alpha\}$ the dual basis, i.e., the 1-forms that satisfy

$$i_\alpha(\omega^\beta) = \delta_\alpha^\beta \,. \tag{2.1}$$

The structure coefficients are defined by the equivalent relations

$$[L_\alpha, L_\beta] = L_\alpha L_\beta - L_\beta L_\alpha = F_{\alpha\beta}{}^\gamma L_\gamma \,, \tag{2.2}$$

$$L_\alpha \omega^\beta = -F_{\alpha\gamma}{}^\beta \omega^\gamma, \tag{2.3}$$

$$d\omega^\alpha = -\tfrac{1}{2} F_{\beta\gamma}{}^\alpha \omega^\beta \wedge \omega^\gamma. \tag{2.4}$$

From the Jacobi identities we get the fundamental relation

$$\{L_\alpha F_{\beta\gamma}{}^\rho\}_{\alpha\beta\gamma} = \{F_{\alpha\beta}{}^\eta F_{\eta\gamma}{}^\rho\}_{\alpha\beta\gamma} \,,$$

 0022-2488/79/122613-06$01.00

where $\{\cdot\}_{\alpha\beta\gamma}$ means cyclic summation over the indices α, β, γ and summing over repeated indices is understood.

The coefficients $F_{\alpha\beta}{}^{\gamma}$ determine the geometric structure of $\mathscr{S}$; for this reason they will be called geometric fields; on the other hand, matter is described by introducing collections of real fields defined on $\mathscr{S}$. Finally, the particular form of geometry, i.e., the equations and the transformation laws that the geometric fields must satisfy, as well as the equations of motion and transformation laws of the matter fields, are derived from an action principle[2] of the form

$$\delta \int_S \lambda = 0, \tag{2.5}$$

where S is an arbitrary four-dimensional compact submanifold of $\mathscr{S}$ and λ is a differential 4-form, called the Lagrangian form.

We will limit ourselves to consider theories where λ can be written as

$$\begin{aligned}\lambda &= \lambda^{\mathscr{G}} + \lambda^{\mathscr{M}} \\ &= \frac{1}{4!}\lambda^{\mathscr{G}}_{\alpha\beta\gamma\delta}(F_{\mu\nu}{}^{\rho})\omega^{\alpha}\wedge\omega^{\beta}\wedge\omega^{\gamma}\wedge\omega^{\delta} \\ &\quad + \frac{1}{4!}\lambda^{\mathscr{M}}_{\alpha\beta\gamma\delta}(f, L_{\rho}f)\omega^{\alpha}\wedge\omega^{\beta}\wedge\omega^{\gamma}\wedge\omega^{\delta}.\end{aligned} \tag{2.6}$$

We note that $\lambda^{\mathscr{M}}$ determines the form of the equations of motion and the transformation laws of the matter fields, while the whole λ is needed in order to determine the geometry of $\mathscr{S}$.

If we carry out variations of the fields that vanish on the whole submanifold S, we get the "normal" field equations, i.e.,

$$\begin{aligned}\omega^{\alpha}\wedge\frac{\partial\lambda}{\partial L_{\beta}f} + \omega^{\beta}\wedge\frac{\partial\lambda}{\partial L_{\alpha}f} = 0, \\ \omega^{\alpha}\wedge\frac{\partial\lambda}{\partial F_{\beta\gamma}{}^{\rho}} + \omega^{\beta}\wedge\frac{\partial\lambda}{\partial F_{\alpha\gamma}{}^{\rho}} = 0.\end{aligned} \tag{2.7}$$

From the first set of normal equation we get

$$\frac{\partial\lambda}{\partial L_{\alpha}f} = \omega^{\alpha}\wedge\rho, \tag{2.8}$$

where

$$\rho = \frac{1}{n-3} i_{\alpha}\frac{\partial\lambda}{\partial L_{\alpha}f}. \tag{2.9}$$

Using the second set of normal field equations we see that, after some calculations, it is possible to write the variation of $\lambda^{\mathscr{G}}$ in the simpler form

$$\begin{aligned}&\delta\lambda^{\mathscr{G}}_{\alpha\beta\gamma\delta}\omega^{\alpha}\wedge\omega^{\beta}\wedge\omega^{\gamma}\wedge\omega^{\delta} \\ &\quad = G_{\rho\gamma\delta}\delta F_{\alpha\beta}{}^{\rho}\omega^{\alpha}\wedge\omega^{\beta}\wedge\omega^{\gamma}\wedge\omega^{\delta}.\end{aligned} \tag{2.10}$$

Now if we consider variations of the fields that vanish on the boundary of S, we have the "tangential field equations"

$$\begin{aligned}d\rho - \frac{\partial\lambda}{\partial f} = 0, \\ d\gamma_{\alpha} + F_{\alpha\beta}{}^{\rho}\omega^{\beta}\wedge\gamma_{\rho} + i_{\alpha}\lambda^{\mathscr{G}} = \tau_{\alpha}{}^{\mathscr{M}},\end{aligned} \tag{2.11}$$

where

$$\gamma_{\alpha} = -\tfrac{1}{12}G_{\alpha\beta\rho}\omega^{\beta}\wedge\omega^{\rho}, \tag{2.12}$$

$$\tau_{\alpha}{}^{\mathscr{M}} = (L_{\alpha}f)p - i_{\alpha}\lambda^{\mathscr{M}}. \tag{2.13}$$

We recall that $\tau_{\alpha}{}^{\mathscr{M}}$ are nothing but the densities and flows of 4-momentum, relativistic angular momentum, and charge of matter; in agreement with ordinary classical field theory, we see that these quantities play the role of sources in Eqs. (2.11), determining the geometric fields $F_{\alpha\beta}{}^{\gamma}$.

As shown by Toller,[2] the differential 3-form

$$\tau_{\alpha} = d\gamma_{\alpha} = \tau_{\alpha}{}^{\mathscr{M}} - F_{\alpha\beta}{}^{\rho}\omega^{\beta}\wedge\gamma_{\rho} - i_{\alpha}\lambda^{\mathscr{G}} \tag{2.14}$$

describe the densities and flows of 4-momentum, relativistic angular momentum, and charge of all the geometric and matter fields; from Eq. (2.14) we see that τ_{α} are exact differential 3-form (generalized Gauss law).

As a direct consequence of the tangential field equations (2.11) we get

$$d(d\gamma_{\alpha}) = d\tau_{\alpha} = 0, \tag{2.15}$$

that is, τ_{α} are conserved quantities.

3. GEOMETRIC FIELDS

Now we are going to discuss a Lagrangian form describing gravitational and Yang–Mills theories in the presence of matter fields. In this section, we focus our attention on the geometric part $\lambda^{\mathscr{G}}$ of the Lagrangian (2.6), while some Lagrangian forms for matter fields will be discussed in the next section.

What we are going to show is the following: Starting from a suitable $\lambda^{\mathscr{G}}$ and assuming a particular form of the matter sources, we shall deduce that

(i) Every point of $\mathscr{S}$ has an open neighborhood which is isomorphic to an open set of the principal fiber bundle of the pseudo-orthogonal reference frames of a space–time manifold.

(ii) It is possible to define a covariant derivative.

(iii) It is possible to determine the transformation properties of the geometric fields with respect to the Lorentz group and the internal symmetry group.

(iv) The theory described by this Lagrangian form is of Einstein–Cartan[7] and Yang–Mills type.

When properties (i) and (ii) hold, we say that we have a theory with a local space–time interpretation.

We emphasize that in the usual theories, where the fields are functions defined on a space–time manifold, the geometric structure of the space–time has to be assumed *a priori*; on the contrary, from the point of view we are considering, they can be derived from a Lagrangian form and an action principle. This is a new interesting feature arising from the ideas developed by Toller.[2]

Now we introduce some conventions and definitions, that will be used in the following. The greek indices $\alpha, \beta, \gamma, \cdots$ will take the values $0,1,\ldots,n-1$; the latin indices $i,j,k,\cdots$ the values 0,1,2,3; the latin indices $a,b,c,\cdots$ will take the values 4,5,...,9; and the capital indices $A,B,C,\cdots$ the values $10,\ldots,n-1$.

Further, g^{ik} will indicate the Minkowski metric tensor, namely

$$g^{00} = 1, \quad g^{11} = g^{22} = g^{33} = -1, \quad g^{ik} = 0 \quad \text{if } i \neq k,$$

and ϵ_{ijrs} will be the usual completely antisymmetric Levi–

Civita symbol normalized to $\epsilon_{0123} = 1$. Finally, we shall denote by $\hat{F}_{\alpha\beta}{}^{\gamma}$ the structure constants of the Lie algebra of the group $\mathscr{P} \times G$, where $\mathscr{P}$ is the Poincaré group and G is a compact semisimple Lie group, the group of internal symmetry. We recall that the only nonvanishing $\hat{F}_{\alpha\beta}{}^{\gamma}$ are

$\hat{F}_{AB}{}^{D}$ structure constants of G;

$\hat{F}_{ab}{}^{d}$ structure constants of the orthochronous Lorentz group;

$\hat{F}_{ai}{}^{k}$ generator of the infinitesimal Lorentz transformations in the 4-vector representation.

An interesting example of a field theory in the space $\mathscr{S}$ of local reference frames was given by Toller;[2] he derived a theory strictly related to Einstein's theory of gravitation and Maxwell's theory of electromagnetism, starting from the following Lagrangian form:

$$\begin{aligned}\lambda^{\mathscr{S}} = {} & \frac{1}{4\kappa} F_{ai}{}^{k} g^{ir} \epsilon_{jkrs} F_{\alpha\beta}{}^{j} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{a} \wedge \omega^{s} \\ & + \frac{1}{4\kappa} F_{ai}{}^{k} g^{ir} \epsilon_{jrpq} F_{bk}{}^{j} \omega^{a} \wedge \omega^{b} \wedge \omega^{p} \wedge \omega^{q} \\ & + \frac{1}{16\pi} F_{ik}{}^{10} g^{ir} g^{ks} F_{rs}{}^{10} \omega^{0} \wedge \omega^{1} \wedge \omega^{2} \wedge \omega^{3} \\ & - \frac{1}{32\pi} F_{ik}{}^{10} g^{ir} g^{ks} \epsilon_{rspq} F_{\alpha\beta}{}^{10} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{p} \wedge \omega^{q}, \end{aligned} \tag{3.1}$$

where κ is the gravitational constant.

We see that this Lagrangian is quadratic in all the geometric fields $F_{\alpha\beta}{}^{\gamma}$; thus gravitation and electromagnetism would be treated in an homogeneous way. Nevertheless, this quadratic Lagrangian is not completely satisfactory, because it does not lead to a complete determination of all the geometric fields $F_{\alpha\beta}{}^{\gamma}$.[8]

In order to avoid this difficulty, we shall consider a Lagrangian form composed by a linear term, proposed by Neéman and Regge,[4] which concerns gravitation, and a quadratic term giving Yang–Mills theory, namely,

$$\begin{aligned}\lambda^{\mathscr{G}} = {} & \frac{1}{8\kappa} \hat{F}_{ai}{}^{k} g^{ir} \epsilon_{krpq} (F_{\alpha\beta}{}^{a} - \hat{F}_{\alpha\beta}{}^{a}) \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{p} \wedge \omega^{q} \\ & + \frac{1}{4\eta^{2}} g_{AB} F_{ik}{}^{A} g^{ir} g^{ks} F_{rs}{}^{B} \omega^{0} \wedge \omega^{1} \wedge \omega^{2} \wedge \omega^{3} \\ & - \frac{1}{8\eta^{2}} g_{AB} F_{ik}{}^{A} g^{ir} g^{ks} \epsilon_{rspq} (F_{\alpha\beta}{}^{B} - \hat{F}_{\alpha\beta}{}^{B}) \\ & \cdot \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{p} \wedge \omega^{q}, \end{aligned} \tag{3.2}$$

where g_{AB} is a nondegenerate invariant quadratic form, i.e., with the property

$$g_{AB} \hat{F}_{CD}{}^{B} + g_{DB} \hat{F}_{CA}{}^{B} = 0, \tag{3.3}$$

and η is the Yang–Mills coupling constant.

Now we are going to derive the equations for the geometric fields. If we perform the variation of the Lagrangian (3.2), we get

$$\begin{aligned}\delta\lambda^{\mathscr{G}}_{\alpha\beta\gamma\delta} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{\gamma} \wedge \omega^{\delta} = {} & \frac{3}{\kappa} \hat{F}_{ai}{}^{k} g^{ir} \epsilon_{krpq} \delta F_{\alpha\beta}{}^{a} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{p} \wedge \omega^{q} \\ & + \frac{12}{\eta^{2}} g_{AB} \delta F_{ik}{}^{A} g^{ir} g^{ks} F_{rs}{}^{B} \omega^{0} \wedge \omega^{1} \wedge \omega^{2} \wedge \omega^{3} \\ & - \frac{3}{\eta^{2}} g_{AB} \delta F_{ik}{}^{A} g^{ir} g^{ks} \epsilon_{rspq} (F_{\alpha\beta}{}^{B} - \hat{F}_{\alpha\beta}{}^{B}) \\ & \times \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{p} \wedge \omega^{q} - \frac{3}{\eta^{2}} g_{AB} F_{ik}{}^{A} g^{ir} g^{ks} \epsilon_{rspq} \\ & \times \delta F_{\alpha\beta}{}^{B} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{p} \wedge \omega^{q}. \end{aligned} \tag{3.4}$$

If the normal equations are satisfied, we can write Eq. (3.4) in the form (2.10), namely,

$$\begin{aligned}\delta\lambda^{\mathscr{G}}_{\alpha\beta\gamma\delta} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{\gamma} \wedge \omega^{\delta} = {} & G_{\rho\gamma\delta} \delta F_{\alpha\beta}{}^{\rho} \omega^{\alpha} \wedge \omega^{\beta} \wedge \omega^{\gamma} \wedge \omega^{\delta}; \end{aligned} \tag{3.5}$$

from Eqs. (3.4) and (3.5) we get the nonvanishing components of $G_{\rho\gamma\delta}$

$$\begin{aligned} G_{apq} = {} & \frac{3}{\kappa} \hat{F}_{ai}{}^{k} g^{ir} \epsilon_{krpq} = -G_{aqp}, \\ G_{Apq} = {} & -\frac{3}{\eta^{2}} g_{AB} F_{ik}{}^{B} g^{ir} g^{ks} \epsilon_{rspq} = -G_{Aqp}, \end{aligned} \tag{3.6}$$

and the conditions

$$\begin{aligned} & F_{b\alpha}{}^{A} = 0, \\ & F_{B\alpha}{}^{A} = \hat{F}_{B\alpha}{}^{A}. \end{aligned} \tag{3.7}$$

From Eqs. (3.6) and (2.12) we find

$$\begin{aligned} & \gamma_{t} = 0, \\ & \gamma_{a} = -\frac{1}{4\kappa} \hat{F}_{ai}{}^{k} g^{ir} \epsilon_{krpq} \omega^{p} \wedge \omega^{q}, \\ & \gamma_{A} = \frac{1}{4\eta^{2}} g_{AB} F_{ik}{}^{B} g^{ir} g^{ks} \epsilon_{rspq} \omega^{p} \wedge \omega^{q}, \end{aligned} \tag{3.8}$$

and after some calculations, using Eqs. (2.11) and (3.7), we have

$$\begin{aligned}\tau_{t}^{\mathscr{M}} = {} & \frac{1}{4\kappa} \hat{F}_{am}{}^{i} g^{mj} \epsilon_{ijtq} \{ (F_{bd}{}^{a} - \hat{F}_{bd}{}^{a}) \omega^{b} \wedge \omega^{d} \wedge \omega^{q} + 2F_{br}{}^{a} \omega^{b} \wedge \omega^{r} \wedge \omega^{q} + 2F_{bB}{}^{a} \omega^{b} \wedge \omega^{B} \wedge \omega^{q} + 2F_{rB}{}^{a} \omega^{r} \wedge \omega^{B} \wedge \omega^{q} \\ & + F_{AB}{}^{a} \omega^{A} \wedge \omega^{B} \wedge \omega^{q} \} + \frac{1}{\kappa} \{ -\hat{F}_{aj}{}^{i} F_{it}{}^{a} g^{jk} + \tfrac{1}{2} \delta_{t}{}^{k} \hat{F}_{aj}{}^{i} F_{ir}{}^{a} g^{jr} \} \sigma_{k} \\ & - \frac{g_{AB}}{\eta^{2}} \{ -F_{ij}{}^{A} F_{ts}{}^{B} g^{js} g^{ik} + \tfrac{1}{4} \delta_{t}{}^{k} F_{ij}{}^{A} F_{rs}{}^{B} g^{js} g^{ir} \} \sigma_{k}, \end{aligned} \tag{3.9}$$

$$\begin{aligned}\tau_{a}^{\mathscr{M}} = {} & \frac{1}{4\kappa} \hat{F}_{am}{}^{i} g^{mj} \epsilon_{ijpq} \{ F_{bd}{}^{p} \omega^{b} \wedge \omega^{d} \wedge \omega^{q} + 2F_{bB}{}^{p} \omega^{b} \wedge \omega^{B} \wedge \omega^{q} + 2F_{rB}{}^{p} \omega^{r} \wedge \omega^{B} \wedge \omega^{q} + 2F_{AB}{}^{p} \omega^{A} \wedge \omega^{B} \wedge \omega^{q} \\ & + 2(F_{br}{}^{p} - \hat{F}_{br}{}^{p}) \omega^{b} \wedge \omega^{r} \wedge \omega^{q} \} + \frac{1}{2\kappa} \hat{F}_{am}{}^{i} g^{mj} \{ F_{ij}{}^{k} + \delta_{i}{}^{k} F_{jr}{}^{r} - \delta_{j}{}^{k} F_{ir}{}^{r} \} \sigma_{k}, \end{aligned} \tag{3.9'}$$

$$\tau_A{}^{\prime\prime} = \frac{1}{4\eta^2} g_{AB}\, g^{ir} g^{ks}\{\epsilon_{rspq} L_b F_{ik}{}^B - 2\epsilon_{rstq} F_{ik}{}^B F_{bp}{}^t\}\omega^b \wedge \omega^p \wedge \omega^q + \frac{1}{4\eta^2}\epsilon_{rspq}\, g^{ir} g^{ks}\{g_{AD} L_B F_{ik}{}^D + F_{ik}{}^E F_{AB}{}^D g_{DE}\}$$
$$\times \omega^B \wedge \omega^p \wedge \omega^q + \frac{1}{\eta^2} g_{AB}\, g^{jm}\{ g^{ik} L_m F_{ij}{}^B - g^{ik} F_{ij}{}^B F_{mr}{}^r - g^{in} F_{ij}{}^B F_{mn}{}^k \}\sigma_k$$
$$+ i_A \lambda^{\prime\prime} + F_{AB}{}^a \omega^B \wedge \gamma_a \,, \tag{3.9''}$$

where σ_k is the differential 3-form given by

$$\sigma_k = \tfrac{1}{6}\epsilon_{kijm}\omega^i \wedge \omega^j \wedge \omega^m.$$

We see that Eqs. (3.9) are very complicated, which have to be satisfied by a set of differential 3-forms.

Nevertheless, if we assume the local form of the sources, namely, if they are of the type

$$\tau_\alpha{}^{\prime\prime} = T_\alpha{}^k \sigma_k \,, \tag{3.10}$$

the system (3.9) can be strongly simplified.

With this assumption we can derive from Eqs. (3.9) other constraints on the geometric fields, i.e.,

$$F_{\alpha A}{}^i = 0, \quad F_{\alpha b}{}^i = \hat{F}_{\alpha b}{}^i \,,$$
$$F_{\alpha A}{}^a = 0, \quad F_{\alpha b}{}^a = \hat{F}_{\alpha b}{}^a \,. \tag{3.11}$$

From Eqs. (3.11) we get

$$F_{ab}{}^i = F_{aA}{}^i = F_{AB}{}^i = 0; \tag{3.11'}$$

these conditions tell us that the $(N+6)$-dimensional distribution $s \mapsto \text{span}\,(A_a, A_B)$ is involutive; then, from the well known Frobenius theorem,[9] it follows that there is a local system of coordinates $(x_0,...,x_{9+N})$ at $s \in \mathcal{S}$, such that the set of coordinates $(x_4,...,x_{9+N})$ defines the $(N+6)$-dimensional integral manifold of the distribution. The remaining set of coordinates $(x_0,...,x_3)$ can be interpreted as space–time local coordinates.

Furthermore, from Eqs. (3.11), it follows that the conditions

$$L_j F_{ai}{}^k = -F_{aj}{}^b F_{bi}{}^k,$$

which allow us to define a covariant derivative,[2] are trivially satisfied. Taking into account Eqs. (3.7) and (3.11), the definitions of the torsion and curvature tensors become[10]

$$S_{ik}{}^r = -F_{ik}{}^r; \tag{3.12}$$
$$R^i{}_{jkm} = -\hat{F}_{aj}{}^i F_{km}{}^a, \tag{3.13}$$

and the Jacobi identities take the form

$$L_{\bar{a}} F_{ij}{}^k = \hat{F}_{\bar{a}i}{}^r F_{rj}{}^k + \hat{F}_{\bar{a}j}{}^r F_{ir}{}^k - \hat{F}_{\bar{a}r}{}^k F_{ij}{}^r;$$
$$L_{\bar{a}} F_{ij}{}^b = \hat{F}_{\bar{a}i}{}^r F_{rj}{}^b + \hat{F}_{\bar{a}j}{}^r F_{ir}{}^b - \hat{F}_{\bar{a}d}{}^b F_{ij}{}^d; \tag{3.14}$$
$$L_{\bar{a}} F_{ij}{}^B = \hat{F}_{\bar{a}i}{}^r F_{rj}{}^B + \hat{F}_{\bar{a}j}{}^r F_{ir}{}^B - \hat{F}_{\bar{a}D}{}^B F_{ij}{}^D,$$

where $\bar{a}$ runs over the values 4,5,...,$n-1$, and

$$\{L_i F_{jk}{}^B\}_{ijk} = -\{F_{tk}{}^B S_{ij}{}^t\}_{ijk} \,,$$
$$\{L_i R^m{}_{njk}\}_{ijk} = -\{S_{ij}{}^t R^m{}_{ntk}\}_{ijk} \,, \tag{3.15}$$
$$\{L_i S_{jk}{}^m\}_{ijk} = \{-S_{ij}{}^r S_{rk}{}^m + R^m{}_{ijk}\}_{ijk} \,.$$

First, we note that the fields which are not structure constants are reduced to $F_{ij}{}^k$, linked to torsion tensor, $F_{ij}{}^a$ to the curvature tensor and $F_{ij}{}^A$ to the gauge field strength.

Further, we see that the transformation laws (3.14) of these fields and the well-known Bianchi identities (3.15)[10] for the torsion, curvature and $F_{ij}{}^A$, and nothing but particular Jacobi identities; we remark that from Eq. (3.14) and definition (3.13) we get the correct homogeneous transformation law for the Riemanian curvature tensor.

Finally, after straightforward calculations, taking into account the constraints of the geometry, the system (3.9) becomes

$$\kappa T_t{}^k = R^i{}_{mit}\, g^{mk} - \tfrac{1}{2}\delta_t{}^k R^i{}_{mij}\, g^{mj} - \frac{\kappa}{\eta^2} g_{AB}$$
$$\times\{\tfrac{1}{4}\delta_t{}^k F_{ij}{}^A F_{rs}{}^B g^{ir} g^{js} - F_{ij}{}^A F_{ts}{}^B g^{ik} g^{js}\} \,, \tag{3.16}$$
$$\kappa T_a{}^k = \tfrac{1}{2} F_{am}{}^i g^{mj}\{S_{ji}{}^k + \delta_j{}^k S_{ir}{}^r - \delta_i{}^k S_{jr}{}^r\} \,, \tag{3.17}$$
$$\eta^2 T_A{}^k = g_{AB}\, g^{ik} g^{jm}\{L_m F_{ij}{}^B + F_{ij}{}^B S_{mr}{}^r\}$$
$$+ \tfrac{1}{2} g_{AB}\, g^{ir} g^{js} F_{ij}{}^B S_{rs}{}^k. \tag{3.18}$$

It can easily be seen that Eqs. (3.16) and (3.17) are just Einstein–Cartan equations in the presence of the Yang–Mills fields, while Eq. (3.18) is the inhomogeneous Yang–Mills equation in the presence of a gravitational field with a nonvanishing torsion.

We observe that the Yang–Mills curvature tensor $F_{ik}{}^A$ is equal to the coupling constant η times the Yang–Mills strength tensor $G_{ik}{}^A$, which is generally used in the physical literature; in this way Eq. (3.18) can be written in the more familiar form

$$\eta T_A{}^k = g_{AB}\, g^{ik} g^{jm}\{L_m G_{ij}{}^B + G_{ij}{}^B S_{mr}{}^r$$
$$+ \tfrac{1}{2} g_{AB}\, g^{ir} g^{js} G_{ij}{}^B S_{rs}{}^k.$$

4. MATTER FIELDS

Here we shall consider some Lagrangian forms concerning the matter fields in the presence of gravitational and Yang–Mills fields, as described by Eq. (3.2) in the preceding section. We limit ourselves to a discussion of the scalar and Dirac fields, but the formalism can easily be extended to higher spins. We will show that, generalizing in a very natural way the results given by Toller and Vanzo,[3] we obtain the correct equations of motion and transformation laws of the matter fields.

We introduce the notation

$$\sigma = \omega^0 \wedge \omega^1 \wedge \omega^2 \wedge \omega^3,$$
$$\sigma_k = i_k \sigma = \tfrac{1}{6}\epsilon_{kijm}\omega^i \wedge \omega^j \wedge \omega^m, \tag{4.1}$$

and we note that

$$\omega^k \wedge \sigma_i = \delta_i{}^k \sigma \,, \tag{4.2}$$
$$d\sigma_i = \tfrac{1}{4}\epsilon_{irst} F_{\alpha\beta}{}^s \omega^\alpha \wedge \omega^\beta \wedge \omega^r \wedge \omega^t. \tag{4.3}$$

Let us consider the following Lagrangian, describing the scalar field:

$$\lambda^{\mathscr{M}} = \tfrac{1}{2}\, g^{ik}\,[(L_i\varphi)^T L_k\varphi - m^2\varphi^T\varphi\,]\sigma + \tfrac{1}{2}\, g^{ik}\,[(L_i\varphi)^T L_a\varphi + (L_a\varphi)^T L_i\varphi\,] \times\omega^a\wedge\sigma_k + \tfrac{1}{2}\, g^{ik}\{\,[(L_A + K_A)\varphi\,]^T L_i\varphi + (L_i\varphi)^T(L_A + K_A)\varphi\,\}\omega^A\wedge\sigma_k\,, \tag{4.4}$$

where φ is a column vector of p-real fields defined on $\mathscr{S}$, and φ^T is the corresponding transpose row vector. The matrices K_A form a real representation of the Lie algebra of G in a p-dimensional vector space V.

The normal field equations (2.7) give

$$L_a\varphi = 0, \quad L_A\varphi = -K_A\varphi. \tag{4.5}$$

These equations show that φ are scalar fields with respect to homogeneous Lorentz transformations and that they just form a multiplet belonging to the representation of the group G in the linear space V.

Taking into account normal equations and Eq. (4.3), the tangential field equations yield

$$g^{ik}L_iL_k\varphi + g^{ik}L_k\varphi S_{ij}{}^j + m^2\varphi = 0, \tag{4.6}$$

$$\tfrac{1}{4}\, g^{ik}\epsilon_{irst}L_k\varphi\,(F_{ab}{}^s\omega^a\wedge\omega^b\wedge\omega^r\wedge\omega^t + 2F_{aB}{}^s\omega^a\wedge\omega^B\wedge\omega^r\wedge\omega^t + F_{AB}{}^s\omega^A\wedge\omega^B\wedge\omega^r\wedge\omega^t) = 0. \tag{4.6'}$$

From the definition (2.13) and the normal equation we have

$$\tau_i{}^{\mathscr{M}} = \tfrac{1}{2}\{\, g^{jk}\,[(L_i\varphi)^T L_k\varphi + (L_k\varphi)^T L_i\varphi] - [\, g^{mk}(L_m\varphi)^T L_k\varphi - m^2\varphi^T\varphi\,]\delta_i{}^j\}\sigma_j\,, \tag{4.7}$$

$$\tau_a{}^{\mathscr{M}} = 0,$$

$$\tau_A{}^{\mathscr{M}} = -\tfrac{1}{2} g^{jk}\,[(L_k\varphi)^T K_A\varphi + (K_A\varphi)^T L_k\varphi\,]\sigma_j\,.$$

These formulas show that we are dealing with local sources; thus Eqs. (3.11) ensure us that (4.6') is identically satisfied.

Since $\tau_a{}^{\mathscr{M}} = 0$, the field equations (3.17) yield

$$S_{jk}{}^i + \delta_j{}^i S_{km}{}^m - \delta_k{}^i S_{jm}{}^m = 0; \tag{4.8}$$

this gives $S_{ij}{}^k = 0$; thus Eq. (4.7) becomes

$$g^{ik}L_iL_k\varphi + m^2\varphi = 0. \tag{4.9}$$

This result shows, as expected, that a scalar field cannot create torsion.

Let us consider the Dirac field. We start from the following Lagrangian:

$$\lambda^{\mathscr{M}} = \tfrac{1}{2}\,[\psi^T\gamma^0\gamma^k\mathscr{R}L_k\psi - (L_k\varphi)^T\gamma^0\gamma^k\mathscr{R}\psi + im\psi^T\gamma^0\mathscr{R}\psi]\sigma + \tfrac{1}{2}\{\psi^T\gamma^0\gamma^i\mathscr{R}(L_a + S_a)\psi - [(L_a + S_a)\psi]^T\gamma^0\gamma^i\mathscr{R}\psi\}\omega^a\wedge\sigma_i + \tfrac{1}{2}\{\psi^T\gamma^0\gamma^i\mathscr{R}(L_A + S_A)\psi - [(L_A + S_A)\psi]^T\gamma^0\gamma^i\mathscr{R}\psi\}\omega^A\wedge\sigma_i\,, \tag{4.10}$$

where ψ is a column vector composed by p 4-component real spinor fields, γ^i are the Dirac matrices in the Majorana representation, the matrices S_a are given by

$$S_a = \tfrac{1}{4}\,\hat{F}_{ak}{}^i g_{ij}\Sigma^{jk} = -\gamma^0 S_a{}^T\gamma^0, \qquad \Sigma^{jk} = \tfrac{1}{2}(\gamma^j\gamma^k - \gamma^k\gamma^j)\,, \tag{4.11}$$

and the matrices S_A form a $4p$-dimensional real representation of the Lie algebra of G; finally $\mathscr{R}$ is a $4p\times 4p$ real metric operator with the property[11]

$$\mathscr{R}S_A - S_A\mathscr{R} = 0. \tag{4.12}$$

We have the following commutation relations:

$$[\gamma^i, S_a] = F_{aj}{}^i\gamma^j, \quad [\gamma^i, S_A] = 0; \tag{4.13}$$

from Eq. (2.9) we get

$$\rho = -\tfrac{1}{2}\gamma^0\gamma^i\mathscr{R}\psi\sigma_i\,. \tag{4.14}$$

Since the Lagrangian (4.10) is linear in the fields ψ, the normal field equations (2.7) are automatically satisfied. In order to write the tangential field equations, we note that

$$\frac{\partial\lambda^{\mathscr{M}}}{\partial\psi^T} = \tfrac{1}{2}(\gamma^0\gamma^i\mathscr{R}L_i\psi + im\gamma^0\mathscr{R}\psi)\sigma + \tfrac{1}{2}\,[\gamma^0\gamma^i\mathscr{R}(L_a + S_a)\psi + S_a\gamma^0\gamma^i\mathscr{R}\psi]\omega^a\wedge\sigma_i + \tfrac{1}{2}\,[\gamma^0\gamma^i\mathscr{R}(L_A + S_A)\psi + S_A\gamma^0\gamma^i\mathscr{R}\psi]\omega^A\wedge\sigma_i\,. \tag{4.15}$$

Then, after some calculations, the tangential field equations yield

$$\left(-i\gamma^kL_k + m - \frac{i}{2}\gamma^kS_{kr}{}^r\right)\psi = 0, \tag{4.16}$$

$$L_a\psi = -S_a\psi, \quad L_A\psi = -S_A\psi, \tag{4.17}$$

and a third equation analogous to Eq. (3.6') which will be identically satisfied for the same reasons; as before, $S_{jk}{}^i$ are the components of the torsion tensor.

Equation (4.16) is the Dirac equation in the presence of gravitation with nonvanishing torsion, while (4.17) are the equations which show that ψ transforms like a multiplet of spin-$\frac{1}{2}$ fields with respect to a Lorentz transformation, and that it belongs to a p-dimensional real representation of the internal symmetry group G.

As a consequence of the field equations (4.16) and (4.17), the Lagrangian (4.10) vanishes identically; then, taking into account the definition (2.13), we have

$$\tau_k{}^{\mathscr{M}} = T_k{}^i\sigma_i = \tfrac{1}{2}\,[\psi^T\gamma^0\gamma^i\mathscr{R}L_k\psi - (L_k\psi)^T\gamma^0\gamma^i\mathscr{R}\psi]\sigma_i\,,$$
$$\tau_a{}^{\mathscr{M}} = T_a{}^i\sigma_i = -\tfrac{1}{2}\,\hat{F}_{ak}{}^j g^{ks}T_{js}{}^i\sigma_i\,, \tag{4.18}$$
$$\tau_A{}^{\mathscr{M}} = T_A{}^i\sigma_i = -\psi^T\gamma^0\gamma^iK_A\mathscr{R}\psi\sigma_i\,,$$

where

$$T_{rs}{}^i = \tfrac{1}{4}\,\psi^T\gamma^0(\gamma^i\Sigma_{rs} + \Sigma_{rs}\gamma^i)\mathscr{R}\psi. \tag{4.19}$$

of 4-momentum, relativistic angular momentum, and charge of a multiplet of spinor fields.

We remark that from Eqs. (3.17), (4.18), and (4.19) we can conclude that $\kappa T_{rs}{}^i = S_{rs}{}^i$ and $S_{ik}{}^k = 0$; this is a particular feature of spin 0 and $\frac{1}{2}$ fields, because for higher spin fields the torsion trace $S_{ik}{}^k$ generally does not vanish. In this way the Dirac equation takes the form

$$(-i\gamma^kL_k + m)\psi = 0.$$

We recall that the Lie derivatives L_i correspond to covariant derivatives;[12] in fact they contain implicitly the minimal coupling between matter and geometric fields.

This can be understood if we compute the commutator

$$[L_i, L_j] = F_{ij}{}^kL_k + F_{ij}{}^aL_a + F_{ij}{}^AL_A\,;$$

for example, in the case of the Dirac field we get

$$[L_i, L_j]\psi = -S_{ij}{}^k L_k \psi - \tfrac{1}{4} R_{sij}{}^r g_{rm} \Sigma^{sm} \psi - F_{ij}{}^A K_A \psi ,$$

and we see that torsion, gravitational and Yang–Mills curvature appear in the right-hand side of this equation.

We conclude this section with some remarks about the conservation laws. We recall that[2] if the Lagrangian form for the matter fields is an homogeneous function of the fields and their derivatives (this is our case), the conservation laws (2.15) can be written in the form

$$d\tau_\alpha{}^{\mathscr{M}} = -F_{\alpha\beta}{}^\rho \omega^\beta \wedge \tau_\rho{}^{\mathscr{M}}. \tag{4.20}$$

From these equations, taking into account the local form of $\tau_\alpha{}^{\mathscr{M}}$, we get the following expressions for the conservation laws, when gravitation with nonvanishing torsion is present, namely,

$$\begin{aligned}
&(L_i + S_{im}{}^m) T_k{}^i \\
&\quad = -S_{ik}{}^j T_j{}^i + \tfrac{1}{2} R^m{}_{nki}\, g^{nj} T_{jm}{}^i + F_{ik}{}^A T_A{}^i , \\
&(L_i + S_{im}{}^m) T_{jk}{}^i = T_{kj} - T_{jk} , \\
&(L_i + S_{im}{}^m) T_A{}^i = 0.
\end{aligned} \tag{4.21}$$

Further, from the same Eq. (4.20), we get the transformation laws of the densities, i.e.,

$$L_a T_j{}^i = \hat{F}_{aj}{}^k T_k{}^i - \hat{F}_{ak}{}^i T_j{}^k ,$$

$$L_a T_{jk}{}^i = \hat{F}_{aj}{}^m T_{mk}{}^i + \hat{F}_{ak}{}^m T_{jm}{}^i - \hat{F}_{am}{}^i T_{jk}{}^m , \tag{4.22}$$

$$L_a T_A{}^i = -\hat{F}_{aj}{}^i T_A{}^j ,$$

$$L_A T_B{}^i = -\hat{F}_{AB}{}^D T_D{}^i . \tag{4.23}$$

Equations (4.22) show that $T_j{}^i$, $T_{jk}{}^i$, $T_A{}^i$ are tensors with respect to a Lorentz transformation, while Eq. (4.23) tells us that the current vector $T_A{}^i$ transforms according to the adjoint representation of G.

We note that, if we are dealing only with scalar fields, we obtain

$$S_{jk}{}^i = 0, \quad T_{jk}{}^i = 0; \tag{4.24}$$

thus, taking into account Eqs. (3.15a) and (3.18), the conservation laws take the more familiar form

$$\begin{aligned}
&L_i (T_k{}^i + T_k{}^{i(YM)}) = 0, \\
&L_i T_A{}^i = 0,
\end{aligned} \tag{4.25}$$

where $T_k{}^{i(YM)}$ are the components of the Yang–Mills energy–momentum tensor, namely,

$$\begin{aligned}
T_k{}^{i(YM)} = \frac{1}{\eta^2} g_{AB} \big(&\tfrac{1}{4} \delta_k{}^i F_{pq}{}^A F_{rs}{}^B g^{pr} g^{qs} \\
&- F_{rs}{}^A F_{kj}{}^B g^{ir} g^{js} \big) .
\end{aligned}$$

Finally, for the Dirac field, we have

$$S_{im}{}^m = 0, \quad \kappa T_{jk}{}^i = S_{jk}{}^i;$$

in this way the conservations laws become

$$\begin{aligned}
L_i (T_k{}^i + T_k{}^{i(YM)}) = S_{ki}{}^j T_j{}^i + \frac{1}{\kappa} R^m{}_{kin}\, g^{jn} S_{jm}{}^i \\
+ \frac{1}{\eta^2} g_{AB}\, g^{jm} g^{ns} F_{mn}{}^A F_{rs}{}^B S_{kj}{}^r ,
\end{aligned}$$

$$L_i T_{jk}{}^i = T_{kj} - T_{jk} ,$$

$$L_i T_A{}^i = 0,$$

where Eqs. (3.15a) and (3.18) have been used again.

5. CONCLUSION

In this paper we have seen that gauge theories in the presence of gravitation can be reformulated in the language of the space of the reference frames in a very natural way; this can be done by generalizing the definition of a local reference frame, i.e., by enlarging the dimensions of the manifold.

Furthermore, the particular structure of the geometry of $\mathscr{S}$ has been derived from a suitable Lagrangian and an action principle; ordinary gauge theories show in this language their local character, but we remark that the formalism could also be employed in describing nonlocal theories.

Finally, we note that the interest in considering gauge theories in the presence of gravitation can be argued if one thinks of a possible "true" unification of electromagnetic, weak and strong interactions.

For example, Georgi and Glashow[13] have shown in their attempt based on an SU(5) gauge group that, in order to obtain the well-known experimental limit of the stability of the proton, it is necessary to introduce in the theory masses of the order of 10^{16} Gev; in this case, gravitation can no longer be neglected.

ACKNOWLEDGMENT

It is a great pleasure to thank M. Toller for continuous helpful discussions and suggestions.

[1] F. Lurçat, Physica **1**, 95 (1964).
[2] M. Toller, Nuovo Cimento B **44**, 67 (1978).
[3] M. Toller and L. Vanzo, Lett. Nuovo Cimento **22**, 345 (1978).
[4] Y. Ne'eman and T. Regge, Phys. Lett. B **74**, 54 (1978). Y. Ne'eman and T. Regge, Rivista Nuovo Cimento **1**, 1 (1978).
[5] See, for example, W. Pauli, *Teoria della Relativita* (Boringhieri, Torino, 1958).
[6] See, for example, Y.M. Cho, J. Math. Phys. **16**, 2029 (1975), and references therein.
[7] F.W. Hehl, P. Von der Heyde, G.D. Kerlick, and J.M. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[8] For further developments, see M. Toller, "Homogeneous Theories of Gravitation in the Space of Reference Frames," University of Trento, Preprint UTF 35 (1978).
[9] C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, 1946).
[10] Y. Choquet-Bruhat, *Géométrie differentielle et systémes extérieurs* (Dunod, Paris, 1969).
[11] We note that $\mathscr{H}$ has to be antisymmetric if ψ is a classical spinor field, while it must be symmetric if we are dealing with anticommuting components of ψ.
[12] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969).
[13] H. Georgi and S.L. Glashow, Phys. Rev. Lett. **32**, 430 (1974).

A first order system of differential equations for covariant σ models

C. Reina
Istituto di Fisica dell'Università, Milano, Italy

M. Martellini[a] and P. Sodano[b]
Department of Physics, University of Alberta, Edmonton T6G 2J1 Canada

(Received 22 May 1979; accepted for publication 21 August 1979)

A first order system of differential equations is obtained for a covariant σ model defined on a Riemannian (or pseudo-Riemannian) manifold of arbitrary dimension n. In the case of compact groups, i.e., SO(n), the first order system coincides with the one yielded by topological arguments. Our considerations hold true also for SO(p,q), $p+q=n+1$, invariance groups. The scale invariance of the problem is discussed.

1. INTRODUCTION

It is well known that the solutions for the SO(3) nonlinear σ model found by Belavin and Polyakov through topological arguments can also be derived by means of purely local tools.[1-3] The same local approach yields a class of solutions even for the noncompact problem with the same dimensionality, i.e., for the SO(2,1) invariant problem defined by R^2. In fact, parametrizing the field variables in complex stereographic coordinates, the action becomes

$$S=\int \frac{\xi_z\bar{\xi}_{\bar{z}}+\xi_{\bar{z}}\bar{\xi}_z}{(1+a\xi\bar{\xi})}\,dz\Lambda\,d\bar{z}, \tag{1}$$

where $a=+1$ for SO(3) and $a=-1$ for SO(2,1). Writing the field equations as the divergence of the energy–momentum tensor, one has the first order system

$$(1+a\xi\bar{\xi})\xi_z\bar{\xi}_z=f(z), \tag{2}$$

where $f(z)$ is analytic. For the compact case ($a=1$), regularity and the finite action condition amount to imposing that $f(z)=0$ (see Ref. 2) and therefore the system (2) reduces to the Cauchy–Riemann equation for the field ξ. The noncompact case ($a=-1$) is equivalent to the problem of finding axisymmetric solutions of the vacuum Einstein equations which are of the Petrov type N.[4] In this case there is no physical reason for the action to be finite. Nevertheless, one can show that the only relevant solutions are still given by $\xi_{\bar{z}}=0$ as in the compact case.

The aim of this paper is to show that, independently of the compactedness of the group and the dimension of the base space, it is possible to find a first order system (f.o.s) naturally associated to a certain class of σ models.

In the next section we obtain the general form of the f.o.s. for a generally covariant σ model defined on a Riemannian or pseudo-Riemannian manifold of arbitrary dimension. The third section is devoted to the analysis of our result in the physically interesting case of $n=4$.

2. GENERALIZED σ MODELS

In this section we study a generally covariant σ model defined on a Riemannian (or pseudo-Riemannian) manifold M of arbitrary dimension n. The invariance group of the model is taken to be SO(p,q) with $p+q=n+1$. Let Φ be a map

$$\Phi: M\to R^{p+q} \tag{3}$$

invariant under the action of SO(p,q) on R^{p+q}. The action of the problem is

$$S=\int *\,d\Phi^a\Lambda\,d\Phi_a=\int\lambda \tag{4}$$

with the constraint

$$\Phi^a\Phi_a=1, \tag{5}$$

where the internal indices are saturated with the Killing form of SO(p,q), d is the exterior differential, Λ is the wedge product, and $*$ is the Hodge duality operator.

Since the action is a geometrical object on the manifold M, it is manifestly covariant under the general group GL(n,R). By varying the action (4) and taking into account the condition (5), one has the following field equations:

$$d*d\Phi_a+\Phi_a(d\Phi^b\Lambda *d\Phi_b)=0. \tag{6}$$

In the case of compact invariance groups (i.e., when $q=0$ and $\Phi: M\to S^n$) it is relevant to consider the following n form:

$$*\rho=\epsilon_{a_1a_2\cdots a_{n+1}}\Phi^{a_1}\,d\Phi^{a_2}\Lambda\cdots\Lambda\,d\Phi^{a_{n+1}}, \tag{7}$$

which upon integration on M gives the homotopy class of the map Φ:

$$Q=\int *\rho\in\Pi_n(M,S^n). \tag{8}$$

For instance, when $n=2$ and $M=S^2$, one has that $Q\in\Pi_2(S^2)=Z$ is the "topological charge" introduced by Belavin and Polyakov. In this case one has also the remarkable inequality $S\geqslant Q$. The condition

$$*d\Phi_{b_1}=-\epsilon_{b_1b_2b_3}\Phi^{b_2}\,d\Phi^{b_3} \tag{9}$$

yields $S=Q$ and therefore it is sufficient for satisfying the field equations. It is also necessary if one restricts oneself to search for finite action solutions. Note that Eq. (9) yields $\lambda=*\rho$. In the noncompact case the topological argument given above does not apply any more. Nevertheless there is a local reason why the condition $\lambda=*\rho$, which yields the f.o.s.

[a] On leave of absence from Instituto Nazionale di Fisica Nucleare, Sezione di Pavia (Italy).
[b] Permanent address: Instituto di Fisica dell'Universitá di Salerno, Baronissi, Salerno (Italy).

 0022-2488/79/122619-02$01.00

$$*d\Phi_{b_1} = -\epsilon_{b_1 b_2 b_3 \cdots b_{n+1}} \Phi^{b_2} d\Phi^{b_3} \Lambda \cdots \Lambda\, d\Phi^{b_{n+1}}, \quad (10)$$

is still relevant from the mathematical point of view, since every solution (if any) of the system (10) is also a solution of the field equations (6). Of course, the converse may not be true. The proof is given in two steps. First, note that for a field satisfying condition (10) one has the identity

$$d\Phi^{b_1} \Lambda \cdots \Lambda\, d\Phi^{b_{n+1}} = \frac{(-1)^q}{(n-1)!} \epsilon^{a_1 a_2 b_1 \cdots b_{n+1}} \Phi_{a_1} *d\Phi_{a_2}, \quad (11)$$

where q is the number of negative eigenvalues of the Killing form of SO(p,q). The second step amounts to substituting Eq. (10) into the field equations. The Laplacian of Φ reads

$$d* d\Phi_{b_1} = -\epsilon_{b_1 b_2 \cdots b_{n+1}} d\Phi^{b_2} \Lambda \cdots \Lambda\, d\Phi^{b_{n+1}}. \quad (12)$$

Using Eq. (11), one has that

$$\begin{aligned} d*d\Phi_{b_1} &= -\delta^{a_1 a_2}_{b_1 b_2} \Phi_{a_1} d\Phi^{b_2} \Lambda *d\Phi_{a_2} \\ &= -\Phi_{b_1} (d\Phi^{b} \Lambda *d\Phi_b), \end{aligned} \quad (13)$$

showing that the field equations are satisfied.

Let us end this section noting that the f.o.s. (10) is scale invariant whenever the action (4) is. It is well known that this is equivalent to the nondimensionality of the constraint (5), which in turn implies a constraint on the dimension n of the basic space. In fact, assuming that the background covariant metric tensor has dimension $n-2$ as in general relativity, one has

$$\begin{cases} \dim[d\Phi^a] = \dim[\Phi^a], \\ \\ \dim[*d\Phi^a] = \left(\dfrac{n^2}{2} - 3n + 4\right) + \dim[\Phi^a]. \end{cases} \quad (14)$$

Accordingly, if $\dim[\Phi^a] = 0$, the action is nondimensional only for $n = 2,4$.

3. CLOSING REMARKS

The construction given above naturally introduces a f.o.s. of differential equations as a sufficient condition for the solution of the second order system (6). It would be very interesting to ascertain the hypotheses restricting the class of fields for which Eq. (10) is also a necessary condition for the solution of the field equations. Note that for $n = 2$ solutions of Eq (10) have been already found both for the compact[1] and the noncompact case.[5] From the physical point of view, however, the only interesting case is a four dimensional covariant problem, with group SO(5) or SO(4,1). De Alfaro, Fubini, and Furlan[6] first studied such models, coupling the Φ field with the gravitational field in a generally invariant fashion. In their approach the gravitational field is a dynamical variable coupled to the energy–momentum tensor of the Φ's through the Einstein equations. Here the metric field is considered as a background field, and no corrections to the curvature due to the presence of the Φ's is introduced.

As an example, for the SO(5) σ model on the instanton background given by[6]

$$ds^2 = \frac{\alpha^2}{(\alpha + \chi^2)^2} \delta_{\mu\nu}\, dx^\mu\, dx^\nu,$$

one finds that

$$\Phi_\mu = \frac{2\alpha\chi_\mu}{\alpha^2 + \chi^2}, \quad \Phi_5 = \frac{\alpha^2 - \chi^2}{\alpha^2 + \chi^2}$$

satisfies the f.o.s.

ACKNOWLEDGMENTS

One of the authors (C.R.) acknowledges the University of Alberta for hospitality. Professor F. De Felice, Professor I. Hauser, and Professor W. Israel are warmly thanked for fruitful discussions.

[1] A.A. Belavin and A.M. Polyakov, JETP Lett. **22**, 246 (1975).
[2] G. Woo, J. Math. Phys. **18**, 1264 (1977).
[3] L.D. Fadeev and V.E. Korepin, Phys. Rep. C **42**, 1 (1978).
[4] U. Chiminazzo, Thesis, University of Milan (1979) (unpublished).
[5] R.B. Hoffman, J. Math. Phys. **10**, 953 (1969).
[6] V. de Alfaro, S. Fubini, and G. Furlan, "Gauge Theories and Strong Gravity," Ref. TH. 2584-CERN.

Internal Lorentz basis for two-particle states[a)]

M. Daumens[b)] and M. Perroud
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Canada

(Received 31 October 1978)

A basis for the space of states of two relativistic particles with arbitrary spins is constructed, suitable for obtaining two variable relativistic expansions of two-body scattering amplitudes. This basis is obtained starting from a representation of an "internal Lorentz group" which acts on the space of the two particle states in the pseudo-barycentric frame ($\mathbf{p}_1/m_1 + \mathbf{p}_2/m_2 = 0$). The "internal Lorentz group" is defined as the maximal subgroup of the direct product of two Lorentz groups, which leaves invariant the null space of $\mathbf{p}_1/m_1 + \mathbf{p}_2/m_2$.

I. INTRODUCTION

The aim of this paper is to construct two-particle states suitable for deriving two-variable relativistic expansions of two-body scattering amplitudes for particles with arbitrary spins. The two variables are the invariant mass and the momentum transfer squared or, in the center of mass frame, the total energy and the scattering angle.

Two variable Lorentz group expansions have been proposed from a different point of view, mainly for zero-spin particles, in a series of papers, reviewed in Ref. 1.

In a previous paper,[2] we considered two nonrelativistic particles. Starting from a barycentric decomposition of the positions and the momenta observables of the two particles, we have defined an "internal Galilei group." It is a subgroup of the direct product of the two Galilei groups and acts only on the internal variables. We have also constructed the "internal Galilei basis" for the two particles in their center of mass frame (where $\mathbf{p}_1 + \mathbf{p}_2 = 0$) through an adequate reducible representation of this internal Galilei group. A basis for the two particles in any frame was deduced by the action of the Galilei boosts of the "diagonal group" (acting on both particles in the same manner). By construction the basis states transform irreducibly under the internal group, and they have the same transformation properties under the diagnonal homogeneous Galilei group as the canonical two-particle basis (which transform irreducibly under the diagonal inhomogeneous Galilei group). The "internal Galilei states" we have constructed have neither a given internal energy nor a given internal momentum, and it is precisely this fact which makes possible a two-variable expansion.

For two relativistic particles with different masses m_1 and m_2, there is no barycentric decomposition, but we can perform a pseudo-barycentric decomposition by considering the quotient of the momenta by the masses. Then in the frame where $\mathbf{p}_1/m_1 + \mathbf{p}_2/m_2 = 0$, as for the nonrelativistic case, we can define an internal Lorentz group and deduce the corresponding Lorentz basis, by using only group theoretical arguments. Unfortunately, this frame is not very useful for applications. The center of mass frame is the most interesting, in particular for obtaining the relationship with the Poincaré irreducible two-particle states. Thus from the previous basis, we construct in the center of mass frame, an internal Lorentz basis, but we lose the explicit action of the internal group on these states. Then by action of the Lorentz boosts we obtain an internal basis for two particles in any frame. By construction these basis states transform under the Lorentz group as the Poincaré irreducible states, and they do not have good transformation properties under translations since they are not eigenstates of the invariant mass. They do, however, make it possible to perform two-variable expansions for the scattering amplitudes.

In Sec. II we recall the description of the one-particle states making use of three of the most important subgroups of the Poincaré group, namely the translation group, the three-dimensional Euclidean group and the Lorentz group. Section III is devoted to the definition of the internal Lorentz group and Sec. IV to the step by step construction of the corresponding internal Lorentz basis. Finally in Sec. V we give the connection between our basis and the Poncaré irreducible ones, and we deduce the transformation properties of the former under the diagonal Lorentz group. The Appendix is a collection of results on the unitary representations of the Lorentz group and their Clebsch–Gordon coefficients.

II. ONE PARTICLE STATES

Since the Wigner analysis,[3] each particle of mass m and spin s is associated with a representation (m,s) of the Poincaré group, specified by the two Casimir operators of the group:

$$P^2 = m^2, \qquad W^2 = -m^2 s(s+1),$$

where $W_\mu = \frac{1}{2}\epsilon_\mu{}^{\nu\delta\sigma} M_{\nu\delta} P_\sigma$ is the relativistic spin operator, P_μ being the generators of the translation group and $M_{\mu\nu}$ the generator of the Lorentz group.

The states of the particle can be characterized in different manners, making use of different subgroups of the Poincaré group. Let us review the bases corresponding to three of the most important subgroups:

(i) *The translation group* $\mathcal{T}$: The linear momentum basis states are denoted by $|(ms)\underline{p}\mu\rangle$, where $\underline{p}$ is the eigenvalue of the operator P and μ is a label removing the degeneracy. For the canonical basis[3] $\mu \equiv \sigma$ is the eigenvalue of $-W_3/p_0$ and for the helicity basis[4] $\mu \equiv \lambda$ is the eigenvalue of

[a)]Work supported by the Echanges France-Québec and by the Ministère de l'Education du Gouvernement du Québec.

[b)]Permanent address: Laboratoire de Physique Théorique, Université de Bordeaux I, Chemin du Solarium 33170 Gradignan, France.

 0022-2488/79/122621-07$01.00

$-W_0/|\mathbf{p}|$. In a reference tetrad $\{\underline{e}_\alpha\}$, the spherical coordinates (a,θ,φ) of the 4-momentum $\underline{p}$ are defined by

$$\underline{p} = m[\cosh a\, \underline{e}_0 + \sinh a\, \underline{u}(\theta,\varphi)], \tag{1a}$$

$$\underline{u}(\theta,\varphi) = \sin\theta\cos\varphi \underline{e}_1 + \sin\theta\sin\varphi \underline{e}_2 + \cos\theta \underline{e}_3, \tag{1b}$$

with $a \in [0,\infty)$, $\theta \in [0,\pi]$, $\varphi \in [0,2\pi)$. The relation between the helicity and canonical basis is

$$|(ms)\underline{p}\lambda\rangle = \sum_{\sigma=-s}^{+s} |(ms)\underline{p}\sigma\rangle D^{s}_{\sigma\lambda}(R_p), \quad R_p = R(\varphi,\theta,0), \tag{2}$$

where $D^{s}_{\sigma\lambda}$ is a Wigner D function.

(ii) *The three-dimensional Euclidean group* $\mathscr{E}(3)$: The angular momentum basis[5] is specified by the chain of groups $\mathscr{P} \supset \mathscr{E}(3) \supset SO(3) \supset SO(2)$. The operators $\mathbf{p}$, $(\mathbf{J})_i = \frac{1}{2}\epsilon_{0ijk} \times M^{jk}$ are generators of $\mathscr{E}(3)$. The states are denoted by $|(ms)p\lambda jn\rangle$, where the modulus p of the 3-momentum and the helicity λ are determined by the Casimir operators of $\mathscr{E}(3)$:

$$\mathbf{p}^2 = p^2, \quad \mathbf{J}\cdot\mathbf{p} = p\lambda, \tag{3}$$

and, of course, $j(j+1)$ and n are eigenvalues of $\mathbf{J}^2$ and J_3. These states are related to the helicity states by

$$|(ms)\underline{p}\lambda\rangle = \sum_{j=|\lambda|}^{\infty} \sum_{n=-j}^{+j} |(ms)p\lambda jn\rangle D^{j}_{n\lambda}(R_p). \tag{4}$$

(iii) *The homogeneous Lorentz group SO* (3,1): The Lorentz basis[6,7] corresponds to the chain of groups $\mathscr{P} \supset SO(3,1) \supset SO(3) \supset SO(2)$. Here $\mathbf{J}$, $(\mathbf{K})_i = M_{0i}$ are the generators of SO(3,1). The states are denoted by $|(ms)\rho\nu jn\rangle$, where $\rho \in [0, +\infty)$ and $\nu \in \{-s, -s+1 \ldots, +s\}$ are related to the eigenvalues of the Casimir operators of SO(3,1)

$$\mathbf{K}^2 - \mathbf{J}^2 = \rho^2 - \nu^2 + 1, \quad \mathbf{K}\cdot\mathbf{J} = \rho\nu. \tag{5}$$

In order to relate this basis to the previous ones, let us consider the following transformations:

$$L_p \equiv L(a,\theta,\varphi) = R(\varphi,\theta,0)B_3(a)R^{-1}(\varphi,\theta,0), \tag{6a}$$

$$\Lambda_p \equiv \Lambda(a,\theta,\varphi) = R(\varphi,\theta,0)B_3(a) = L_p R_p, \tag{6b}$$

where $B_3(a)$ is a boost along the 3-axis, and L_p a boost in the direction (θ,φ). Chakrabarti[7] has shown that the Lorentz basis is related to the linear momentum bases by means of the Lorentz unitary representation matrices (see Appendix) $D^{\rho\nu}_{jns\sigma}(L_p)$ or $D^{\rho\nu}_{jns\lambda}(\Lambda_p)$:

—Canonical basis

$$|(ms)\underline{p}\sigma\rangle = \sum_{j=\{0,1/2\}}^{\infty} \sum_{n=-j}^{+j} \sum_{\nu} \int_0^\infty \mu(\rho,\nu)d\rho \times N(\rho,\nu) D^{\rho\nu}_{jns\sigma}(L_p)|(ms)\rho\nu jn\rangle, \tag{7a}$$

—Helicity basis

$$|(ms)\underline{p}\lambda\rangle = \sum_{j=|\lambda|}^{\infty} \sum_{n=-j}^{+j} \sum_{\nu} \int_0^\infty \mu(\rho,\nu)d\rho \times N(\rho,\nu) D^{\rho\nu}_{jns\lambda}(\Lambda_p)|(ms)\rho\nu jn\rangle, \tag{7b}$$

where the integer or half-integer ν varies from $-\min(j,s)$ to $+\min(j,s)$, $\mu(\rho,\nu)$ is the Lorentz invariant measure defined below, and $N(\rho,\nu)$ is a normalization factor. For the two three-parameter families of Lorentz transformations $\{L_p\}$, $\{\Lambda_p\}$, the general completeness and orthogonality relations (A7) and (A8) reduce to

$$\sum_{\sigma} \int \mu(p) d^3p\, D^{\rho\nu}_{jns\sigma}(L_p) D^{\rho'\nu'}_{j'n's\sigma}(L_p)^* = \frac{2\pi^2(2s+1)}{\mu(\rho,\nu)} \delta(\rho-\rho')\delta_{\nu\nu'}\delta_{jj'}\delta_{nn'}, \tag{8}$$

$\mu(p) = \sinh\theta\sinh^2 a$, $d^3p = da\,d\theta\,d\varphi$, $\mu(\rho,\nu) = \rho^2 + \nu^2$,

and

$$\sum_{jn}\sum_{\nu}\int_0^\infty \mu(\rho,\nu)d\rho\, D^{\rho\nu}_{jns\sigma'}(L_p)^* D^{\rho\nu}_{jns\sigma}(L_p) = \frac{2\pi^2(2s+1)}{\mu(p)} \delta^3(p-p')\delta_{\sigma\sigma'}, \tag{9}$$

$\delta^3(p-p') = \delta(a-a')\delta(\theta-\theta')\delta(\varphi-\varphi')$,

with exactly the same relations for the family $\{\Lambda_p\}$. Equation (8) shows that we can invert Eqs. (7) by summing over the magnetic quantum number σ (or λ) and by integrating over the three spherical components (a,θ,φ) of p:

$$|(ms)\rho\nu jn\rangle = [2\pi^2(2s+1)N(\rho,\nu)]^{-1} \times \sum_{\sigma} \int \mu(p) d^3p\, D^{\rho\nu}_{jns\sigma}(L_p)^* |(ms)\underline{p}\sigma\rangle \tag{10a}$$

$$= [2\pi^2(2s+1)N(\rho,\nu)]^{-1} \times \sum_{\lambda} \int \mu(p) d^3p\, D^{\rho\nu}_{jns\lambda}(\Lambda_p)^* |(ms)\underline{p}\lambda\rangle. \tag{10b}$$

The relation of the Lorentz basis to the angular momentum one involves only the boost matrices

$$|(ms)p\lambda jn\rangle = \sum_{\nu}\int_0^\infty d\rho\, N(\rho,\nu) d^{\rho\nu}_{j\lambda s}(a)|(ms)\rho\nu jn\rangle \tag{11}$$

and can be inverted by using the orthonormality relation (A5) of these boost matrices:

$$|(ms)\rho\nu jn\rangle = 2[\pi(2s+1)(2j+1)N(\rho,\nu)]^{-1} \times \sum_{\lambda}\int_0^\infty \sinh^2 a\, da\, d^{\rho\nu}_{j\lambda s}(a)^* |(ms)p\lambda jn\rangle. \tag{12}$$

Note that, in addition to the Poincaré (m,s) labels, all the basis states are characterized by four further variables $\{a,\theta,\varphi,\sigma$ (or $\lambda)\}$, $\{p,\lambda,j,n\}$, $\{\rho,\nu,j,n\}$. For fixed $p = \sinh a$ and λ, Eq. (4) shows that both couples (θ,φ) and (j,n) correspond to each other, while, for fixed j and n, Eqs. (8) and (12) show that both couples $(p = \sinh a, \lambda)$ and (ρ,ν) are in correspondence.

III. The "internal Lorentz group" $\tilde{O}(3,1)$ of a two-particle system

Let us consider two particles of masses, spins, and 4-momenta $m_i, s_i, \underline{p}_i$ $(i = 1,2)$. Let $\underline{x}$ be a timelike unit 4-vector on the upper sheet of the hyperboloid $\mathscr{H}^+$:

$$\underline{x}^2 = 1, \quad x^0 > 0. \tag{13}$$

We define the subset $\mathscr{E}$ of pairs $(\underline{p}_1,\underline{p}_2)$ by

$$\mathscr{E} = \{(\underline{p}_1,\underline{p}_2) | \underline{p}_1 = m_1\underline{x},\ \underline{p}_2 = m_2 \Pi \underline{x},\ \forall\ \underline{x} \in \mathscr{H}^+\}, \tag{14}$$

where Π is the parity operation. Now the transformations of $O^1(3,1) \times O^2(3,1)$, which leave the subset $\mathscr{E}$ invariant, can be

expressed as

$$\widetilde{\Lambda} = \Lambda \times \Pi\Lambda\Pi^{-1}. \tag{15a}$$

The set of such transformations $\widetilde{\Lambda}$ forms a group which we shall call the "internal Lorentz group," and denote by $\widetilde{O}(3,1)$. It is conjugate to the "diagonal" group O(3,1)

$$\widetilde{O}(3,1) = \{\widetilde{\Lambda}\} = (I\times\Pi)O(3,1)(1\times\Pi)^{-1}. \tag{15b}$$

If $\mathbf{J}^{(i)},\mathbf{K}^{(i)}$ are the generators of the group $O^{(i)}(3,1)$, then the generators of the group $\widetilde{O}(3,1)$ are

$$\mathbf{J} = \mathbf{J}^{(1)} + \mathbf{J}^{(2)}, \quad \mathbf{A} = \mathbf{K}^{(1)} - \mathbf{K}^{(2)} \tag{16a}$$

while those of the "diagonal" O(3,1) are, of course,

$$\mathbf{J} = \mathbf{J}^{(1)} + \mathbf{J}^{(2)}, \quad \mathbf{K} = \mathbf{K}^{(1)} + \mathbf{K}^{(2)}. \tag{16b}$$

Note that, since parity commutes with rotations, the diagonal O(3,1) and the internal $\widetilde{O}(3,1)$ have the same O(3) subgroup, the generators of which are **J**.

The total 4-momentum $\underline{p}$ and the Poincaré invariant mass w are

$$P = \underline{p}_1 + \underline{p}_2, \quad w = (\underline{p}^2)^{1/2}. \tag{17}$$

We define three unit timelike 4-vectors:

$$\underline{p} = P/w, \tag{18a}$$

$$\underline{q} = (\hat{p}_1 + \hat{p}_2)/((\hat{p}_1 + \hat{p}_2)^2)^{1/2}, \quad \hat{p}_i = \underline{p}_i/m_i, \tag{18b}$$

$$\underline{x} = (\hat{p}_1 + \Pi\hat{p}_2)/((\hat{p}_1 + \Pi\hat{p}_2)^2)^{1/2} \tag{18c}$$

(for equal masses the 4-vectors $\underline{p}$ and $\underline{q}$ coincide).

In the center of mass frame, the spherical coordinates of $\underline{p}_1$ and $\underline{p}_2$ [see Eq. (1)] are chosen to be (a_1,θ,φ), $(-a_2,\theta,\varphi)$, with the condition

$$m_1\sinh a_1 = m_2\sinh a_2, \quad a_i\in[0,\infty). \tag{19}$$

In this frame, the spherical coordinates of the 4-vectors $\underline{x}$ and $\underline{q}$ are (a,θ,φ) and (b,θ,φ) respectively, where a and b are

$$a = (a_1 + a_2)/2, \quad b = (a_1 - a_2)/2, \tag{20a}$$

and Eq. (19) implies that b is a function of $\underline{a}$:

$$\cosh b = \cosh a\left(1 + \frac{4m_1m_2}{(m_1+m_2)^2}\sinh^2 a\right)^{-1/2} \quad (=1 \text{ if } m_1 = m_2). \tag{20b}$$

In the general mass case, we define the $\underline{q}$ frame by the condition $\underline{q} = \underline{e}_0$, and we denote by L_q the pure Lorentz transformation which maps $\underline{p}$ into $\underline{q}$. The transformation L_q relates the CM frame to the $\underline{q}$ frame.

In the $\underline{q}$ frame the 4-vector $\underline{x}$, defined in Eq. (18c), is related to the 4-momenta of the particles by

$$(\underline{p}_1,\underline{p}_2) = (m_1\underline{x}, m_2\Pi\underline{x})\in\mathscr{E}. \tag{21}$$

Any pair of $\mathscr{E}$ can be obtained from the pair $(m_1\underline{e}_0, m_2\underline{e}_0)$ (both particles at rest) by means of the transformation $\widetilde{L}_x\in\widetilde{O}(3,1)$, where L_x is the pure transformation which maps $\underline{e}_0$ onto $\underline{x}$:

$$(m_1\underline{x}, m_2\Pi\underline{x}) = \widetilde{L}_x(m_1\underline{e}_0, m_2\underline{e}_0). \tag{22}$$

The internal transformation $\widetilde{L}_x$ changes the invariant mass w of the two-particle system. For both particles at rest $\underline{p} = (m_1 + m_2, \mathbf{0})$, but after the transformation $\widetilde{L}_x$ the total 4-momenta and the invariant mass are

$$\underline{p} = [(m_1 + m_2)\cosh a, (m_1 - m_2)\sinh a\,\mathbf{u}(\theta,\varphi)], \tag{23a}$$

$$w^2 = (m_1 + m_2)^2 + 4m_1m_2\sinh^2 a, \tag{23b}$$

i.e., we have $w\in[m_1 + m_2, \infty)$. It is precisely the operators representing these transformations which provide us with the basis functions for expanding two-particle states with arbitrary invariant mass w.

IV. THE INTERNAL LORENTZ BASIS FOR THE TWO-PARTICLE STATES

First, let us start by considering the two particles in the $\underline{q}$ frame where the internal Lorentz group $\widetilde{O}(3,1)$ is defined. The action of the element $\widetilde{\Lambda}\in\widetilde{O}(3,1)$ on the two-particle states is represented by the unitary operator $\mathscr{V}$ defined by

$$\mathscr{V}(\widetilde{\Lambda}) = \mathscr{U}_1(\Lambda)\times\mathscr{U}_2(\Pi\Lambda\Pi^{-1}). \tag{24}$$

Hence any two-particle state in this frame can be obtained from a state with both particles at rest by means of the operator $\mathscr{V}(\widetilde{L}_x)$,i.e.,

$$|m_1\underline{x}\sigma_1\rangle\otimes|m_2\Pi\underline{x}\sigma_2\rangle = \mathscr{V}(\widetilde{L})|m_1\underline{e}_0\sigma_1\rangle\otimes|m_2\underline{e}_0\sigma_2\rangle. \tag{25}$$

Let us introduce the one-particle Lorentz basis

$$|m_1\underline{x}\sigma_1\rangle = \sum_{j_1n_1}\sum_{\nu_1}\int_0^\infty d\rho_1 N(\rho_1,\nu_1)|\rho_1\nu_1 j_1n_1\rangle \times D^{\rho_1\nu_1}_{j_1n_1s_1\sigma_1}(L_x), \tag{26a}$$

$$|m_2\Pi\underline{x}\sigma_2\rangle = \sum_{j_2n_2}\sum_{\nu_2}\int_0^\infty d\rho_2 N(\rho_2\nu_2)|\rho_2\nu_2 j_2n_2\rangle \times D^{\rho_2\nu_2}_{j_2n_2s_2\sigma_2}(\Pi L_x\Pi). \tag{26b}$$

Then by using the symmetries (A6f) and (A4d) of boost matrices, we deduce the identity

$$D^{\rho\nu}_{j_1m_1j_2m_2}(\Pi\Lambda\Pi) = (-1)^{j_1-j_2}D^{\rho-\nu}_{j_1m_1j_2m_2}(\Lambda). \tag{27}$$

Thus, in the tensor product of states (25) we can reduce the product of representation matrices of the same argument L_x by means of the reduction formula (A9). We get

$$|m_1\underline{x}\sigma_1\rangle\otimes|m_2\Pi\underline{x}\sigma_2\rangle = \sum_s\langle s_1\sigma_1s_2\sigma_2|s\sigma\sum_{jn\nu}\int_0^\infty d\rho\mu(\rho,\nu) \times|(\rho\nu js)\underline{q} = \underline{e}_0n\rangle D^{\rho\nu}_{jns\sigma}(L_x), \tag{28}$$

where the states $|(\rho\nu js)\underline{q} = \underline{e}_0n\rangle$ constitute the internal Lorentz basis of the $\underline{q}$ frame and are defined by

$$|(\rho\nu js)\underline{q} = \underline{e}_0n\rangle = \sum_{\substack{j_1n_1\\ j_2n_2}}\sum_{\nu_1\nu_2}\int\int d\rho_1d\rho_2\, N(\rho_1,\nu_1)N(\rho_2,\nu_2)\begin{bmatrix}\rho_1\nu_1\rho_2-\nu_2 & \rho\nu\\ j_1s_1j_2s_2 & js\end{bmatrix} \times(-1)^{j_2-s_1}\mu(\rho,\nu)\langle j_1n_1j_2n_2|jn\rangle \times|\rho_1\nu_1j_1n_1\rangle\otimes|\rho_2\nu_2j_2n_2\rangle, \tag{29}$$

where the coefficient

$$\begin{bmatrix}\rho_1\nu_1\rho_2-\nu_2 & \rho\nu\\ j_1s_1j_2s_2 & js\end{bmatrix}$$

defined in Eq. (A13) is a product of two $\widetilde{O}(3,1)$ reduced Clebsh–Gordan coefficients. By construction these basis states transform irreducibly under a transformation of the group $\widetilde{O}(3,2)$:

$$\mathscr{V}(\widetilde{\Lambda})|(\rho\nu js)\underline{q}_0 = \underline{e}_0n\rangle$$

$$= \sum_{j'n'} |(\rho\nu j's)\underline{q} = \underline{e}_0 n'\rangle D^{\rho\nu}_{j'n'jn}(\Lambda). \tag{30}$$

Thus ρ and ν are related to the eigenvalues of the Casimir operators of $\tilde{O}(3,1)$ while j and n are the total angular momentum and its projection in the $\underline{q}$ frame. Unfortunately, this frame is not very useful for applications. The center of mass frame is the most interesting one, in particular to obtain simple relations with the Poincaré irreducible two particle states: The spin of the two particle system is the total angular momentum in the CM frame. For equal masses the two frames coincide, but in the general mass case we have to generalize our construction to the CM frame.

In the CM frame any two-particle state can be obtained from the rest state of the two particles by

$$|p_1\sigma_1\rangle \otimes p_2\sigma_2\rangle = \mathscr{U}(L_q^{-1})\mathscr{V}(\tilde{L}_x)|m_1\underline{e}_0\sigma_1\rangle \otimes |m_2\underline{e}_0\sigma_2\rangle, \tag{31}$$

where L_q is a pure Lorentz transformation of the diagonal group which maps the CM frame into the $\underline{q}$ frame. Let us recall that L_q is also a function of $\underline{x}$. Then we have

$$|\underline{p}_1\sigma_1\rangle = \sum_{\substack{j_1n_1\\ j_1'n_1'}} \sum_{\nu_1} \int_0^\infty d\rho_1\, N(\rho_1,\nu_1)|\rho_1\nu_1 j_1 n_1\rangle \times D^{j_1n_1j_1'n_1'}_{\rho_1\nu_1}(L_q^{-1})D^{j_1'n_1's_1\sigma_1}_{\rho_1\nu_1}(L_x), \tag{32a}$$

$$|p_2\sigma_2\rangle = \sum_{\substack{j_2n_2\\ j_2'n_2'}} \sum_{\nu_2} \int_0^\infty d\rho\, N(\rho_1,\nu_1)|\rho_1\nu_1 j_1 n_1\rangle \times D^{\rho_2\nu_2}_{j_2n_2j_2'n_2'}(L_q^{-1})(-1)^{s_2-j_2'}D^{\rho_2-\nu_2}_{j_2'n_2's_2\sigma_2}(L_x). \tag{32b}$$

Now by reducing the two tensor products of O(3,1) representation matrices, we get

$$|\underline{p}_1\sigma_1\rangle \otimes |\underline{p}_2\sigma_2\rangle = \sum_s \langle s_1\sigma_1 s_2\sigma_2|s\sigma\rangle \sum_{\substack{jn\\ j'n'}} \sum_{\nu\nu'} \int\int d\rho d\rho'\, \mu(\rho,\nu)\mu(\rho',\nu') \times |(\rho'\nu'j's)\rho\nu jn\rangle D^{\rho\nu}_{jnj'n'}(L_q^{-1})D^{\rho'\nu'}_{j'n's\sigma}(L_x), \tag{33}$$

where we have defined

$$|(\rho'\nu'j's)\rho\nu jn\rangle = \sum_{\substack{jj_2\\ n_1n_2}} \sum_{\substack{j'j_2'\\ \nu_1\nu_2}} \int\int d\rho\, N(\rho_1,\nu_1)d\rho_2 N(\rho_2,\nu_2) \begin{bmatrix} \rho_1\nu_1\rho_2\nu_2 & \rho\nu \\ j_1j_1'j_2j_2' & jj' \end{bmatrix} \times \mu(\rho',\nu')(-1)^{j_2'-s_2} \begin{bmatrix} \rho_1\nu_1\rho_2-\nu_2 & \rho'\nu' \\ j_1's_1j_2's_2 & j's \end{bmatrix} \times \langle j_1n_1j_2n_2|jn\rangle \otimes |\rho_2\nu_2j_2n_2\rangle. \tag{34}$$

Here ρ',ν' specify the representation of the internal Lorentz group $\tilde{O}(3,1)$ while ρ,ν specify the representation of the diagonal Lorentz group. The numbers s,j', and j are the total angular momentum when the two particles are at rest in the $\underline{q}$ frame and in the CM frame, respectively.

Note that Eq. (33) cannot be inverted because L_q is a function of $\underline{x}$. The product of O(3,1) matrices can be written as

$$\sum_{n'} D^{\rho\nu}_{jnj'n'}(L_q^{-1})D^{\rho'\nu'}_{j'n's\sigma}(L_x) = \sum_\lambda D^j_{n\lambda}(\varphi,\theta,0)d^{\rho\nu}_{jj'\lambda}(-b)d^{\rho'\nu'}_{j's\lambda}(a)D^j_{\sigma\lambda}(\varphi,\theta,0)^*, \tag{35}$$

where b is a function of a given in Eq. (20b). We can again reconstruct an O(3,1) matrice corresponding to a pure Lorentz transformation, by means of the coefficients

$$Q(\rho'\nu'j's\rho\nu j;\rho''\nu'') = \sum_\lambda \int_0^\infty \sinh^2 a da\, d^{\rho''\nu''}_{js\lambda}(a)^* d^{\rho\nu}_{jj'\lambda}(-b)d^{\rho'\nu'}_{j's\lambda}(a). \tag{36}$$

The completeness relation (A6) for the matrices $d^{\rho\nu}_{js\lambda}(a)$ allows to invert Eq. (36), and finally Eq. (35) becomes

$$\sum_{n'} D^{\rho\nu}_{jnj'n'}(L_q^{-1})D^{\rho'\nu'}_{j'n's\sigma}(L_x) = \sum_{\nu''} \int_0^\infty d\rho\, \mu(\rho'',\nu')Q(\rho'\nu'j's\rho\nu j;\rho''\nu'')D^{\rho''\nu''}_{jns\sigma}(L_x). \tag{37}$$

We now define the internal Lorentz basis in the CM frame by

$$|(\rho\nu js)\underline{e}_0 n\rangle = \sum_{j'\nu'\nu''} \int\int d\rho' d\rho''\, Q(\rho'\nu'j's\rho''\nu''j;\rho\nu)|(\rho'\nu'j's)\rho''\nu''jn\rangle \tag{38}$$

which yields the expansion

$$|\underline{p}_1\sigma_1\rangle \otimes |\underline{p}_2\sigma_2\rangle = \sum_s \langle s_1\sigma_1 s_2\sigma_2|s\sigma\rangle \sum_{jn\nu} \int_0^\infty \mu(\rho,\nu)d\rho\, |(\rho\nu js)\underline{e}_0 n\rangle D^{\rho\nu}_{jns\sigma}(L_x). \tag{39}$$

The orthogonality of the O(3) C-G coefficients and Eq. (8) allow us to invert the previous relation:

$$|(\rho\nu js)\underline{e}_0 n\rangle = \frac{1}{2\pi^2(2s+1)} \sum_{\sigma_1\sigma_2\sigma} \langle s_1\sigma_1 s_2\sigma_2|s\sigma\rangle \times \int d^3x\, \mu(x)D^{\rho\nu}_{jns\sigma}(L_x)^*|\underline{p}_1\sigma_1\rangle \otimes |\underline{p}_2\sigma_2\rangle. \tag{40}$$

In all this section we have only used the canonical basis, for the helicity basis we have the same relation with the transformation L_x replaced by Λ_x,

$$|\underline{p}_1\lambda_1\rangle \otimes |p_2\lambda_2\rangle = \sum_s \langle s_1\lambda_1 s_2\lambda_2|s\lambda_0\rangle \sum_{jn\nu} \int_0^\infty \mu(\rho,\nu)d\rho\, |(\rho\nu js)\underline{e}_0 n\rangle D^{\rho\nu}_{jns\lambda}(\Lambda_x), \tag{41}$$

$$|(\rho\nu js)\underline{e}_0 n\rangle = \frac{1}{2\pi^2(2s+1)} \sum_{\lambda_1\lambda_2\lambda} \langle s_1\lambda_1 s_2\lambda_2|s\lambda\rangle \times \int d^3x\mu(x)D^{\rho\nu}_{jns\lambda}(\Lambda_x)^*|\underline{p}_1\lambda_1\rangle \otimes |\underline{p}_2\lambda_2\rangle. \tag{42}$$

V. CONNECTION WITH POINCARÉ IRREDUCIBLE STATES AND TRANSFORMATION PROPERTIES FOR THE DIAGONAL LORENTZ GROUP

In an arbitrary frame, we can fully describe the two-particle states in linear momentum basis by means of the two timelike unit 4-vectors $\underline{p}$ and $\underline{x}$, and we note that

$$|\underline{p}_1\sigma_1\rangle \otimes |\underline{p}_2\sigma_2\rangle \equiv |\underline{p},\underline{x},\sigma_1\sigma_2\rangle \equiv |\underline{p},a\theta\varphi,\sigma_1\sigma_2\rangle, \tag{43a}$$

$$|\underline{p}_1\lambda_1\rangle \otimes |\underline{p}_2\lambda_2\rangle \equiv |\underline{p},\underline{x},\lambda_1\lambda_2\rangle \equiv |\underline{p},a\theta\varphi,\lambda_1\lambda_2\rangle. \tag{43b}$$

The Poincaré irreducible two particle states can be de-

fined to be[8]—Canonical (l-s) coupling

$$|w(a)j\underline{p}n,ls\rangle = \sum_m \sum_{\sigma_1\sigma_2\sigma} \langle lms\sigma|jn\rangle\langle s_1\sigma_1 s_2\sigma_2|s\sigma\rangle \frac{1}{\sqrt{4\pi(2l+1)}} \times \int_{S^2} d\Omega(\theta,\varphi)|\underline{p},a\theta\varphi,\sigma_1\sigma_2\rangle,$$

—Helicity coupling

$$|w(a)j\underline{p}n,\lambda_1\lambda_2\rangle = \frac{1}{4\pi}\int_{S^2} d\Omega(\theta,\varphi)D^j_{n\lambda_1+\lambda_2}(\varphi,\theta,0)^*|\underline{p},a\theta\varphi,\lambda_1\lambda_2\rangle. \quad (44b)$$

Introducing Eqs. (44a) and (44b) into Eqs. and (42) respectively, we obtain the relation between the internal Lorentz basis and the Poncaré irreducible bases in the CM frame:

$$|(\rho\nu js)\underline{e}_0 n\rangle = \frac{2}{\pi(2s+1)(2j+1)}\sum_{l\lambda}(2l+1)\langle l\,0s\lambda\,|\,j\lambda\,\rangle \times \int_0^\infty \sinh^2 a da\, d^{\rho\nu}_{js\lambda}(a)^*|w(a)j\underline{e}_0 n,ls\rangle \quad (45a)$$

$$= \frac{2}{\pi(2s+1)}\sum_{\lambda_1\lambda_2\lambda}\langle s_1\lambda_1 s_2\lambda_2|s\lambda\,\rangle \times \int_0^\infty \sinh^2 a da\, d^{\rho\nu}_{js\lambda}(a)^*|w(a)j\underline{e}_0 n\lambda_1\lambda_2\rangle. \quad (45b)$$

These relations show that j is a Poincaré invariant, the spin of the system, and that the integration over the variable a is in fact an integration over the invariant mass $w(a)$ from the threshold to infinity, See Eq. (23).

The completeness relatio (A6) allows us to invert the previous equations; we get

$$|w(a)j,\underline{e}_0 n,ls\rangle = \frac{1}{2j+1}\sum_\lambda \langle l\,0s\lambda\,|j\lambda\,\rangle \sum_\nu \int_0^\infty \mu(\rho,\nu)d\rho \times d^{\rho\nu}_{js\lambda}(a)|(\rho\nu js)\underline{e}_0 n\rangle, \quad (46a)$$

$$|w(a)j,\underline{e}_0 n,\lambda_1\lambda_2\rangle = \frac{1}{2j+1}\sum_s\langle s_1\lambda_1 s_2\lambda_2|s\lambda\,\rangle \times \sum_\nu\int_0^\infty \mu(\rho,\nu)d\rho d\, d^{\rho\nu}_{js\lambda}(a)|(\rho\nu js)\underline{e}_0 n\rangle. \quad (46b)$$

The dependence on the invariant mass $w(a)$ is now fully exhibited in the O(3,1) matrices $d^{\rho\nu}_{js\lambda}(a)$.

Since they are not eigenstates of the linear momentum squarred p^2, the states $|(\rho\nu js)\underline{e}_0 n\rangle$ do not have good transformation properties under the translations; but Eqs. (45) show that they have the same transformation properties under the diagonal Lorentz group as the Poincaré irreducible states, i.e.,

$$\mathcal{U}(\Lambda)|(\rho\nu js)\underline{e}_0 n\rangle = \sum_{n'}|(\rho\nu js)\Lambda\underline{e}_0 n'\rangle D^j_{n'n}[L^{-1}_{\Lambda\underline{e}_0}\Lambda\,], \quad (47)$$

where the Wigner rotation $L^{-1}_{\Lambda\underline{e}_0}\Lambda$ is an element of the little group of $\underline{e}_0$, which reduces to identity if Λ is a pure Lorentz transformation.

We can now generalize the definition of the internal Lorentz basis to an arbitrary frame by the formula

$$|(\rho\nu js)\underline{p}n\rangle = \mathcal{U}(L_p)|(\rho\nu js)\underline{e}_0 n\rangle, \quad (48)$$

where we must remember that $\underline{p}$ is not the total 4-momentum of the two particles but only its direction, i.e., $\underline{P}/w$. The states $|(\rho\nu js)\underline{p}n\rangle$ are related to the Poincaré irreducible ones by the same relations (45) and (46) in which we must replace $\underline{e}_0$ by $\underline{p}$.

Let us consider the action of the parity operator on the linear momentum one-particle states[3,8]

$$\mathcal{U}(\Pi)|\underline{p}_i\sigma_i\rangle = \epsilon_i|\Pi\underline{p}_i\sigma_i\rangle, \quad i=1,2, \quad (49)$$

where ϵ_1 and ϵ_2 are the intrinsic parities of the two particles. Using the identity $L_{\Pi x} = \Pi L_x \Pi^{-1}$, Eqs. (27) and (40), we deduce the action of the parity operation on the internal Lorentz basis

$$\mathcal{U}(\Pi)|(\rho\nu js)\underline{e}_0 n\rangle = \epsilon_1\epsilon_2(-1)^{j-s}|(\rho-\nu js)\underline{e}_0 n\rangle. \quad (50)$$

The time reversal operator τ is represented on the one particle state by the antiunitary operator[8,9]

$$\mathcal{A}(\tau)|\underline{p}_i\sigma_i\rangle = (-1)^{s_i+\sigma_i}|\Pi\underline{p}_i - \sigma_i\rangle. \quad (51)$$

Using the symmetries of the O(3) CG coefficients and the identity

$$D^{\rho\nu}_{jns\sigma}(\Lambda)^* = (-1)^{n-\sigma}D^{\rho\nu}_{j-ns-\sigma}(\Lambda), \quad (52)$$

we obtain

$$\mathcal{A}(\tau)|(\rho\nu js)\underline{e}_0 n\rangle = (-1)^{j+n}|(\rho-\nu js)\underline{e}_0 - n\rangle. \quad (53)$$

Let us call $\mathcal{E}$ the operator which exchange particles 1 and 2. Under this exchange $\underline{x}$ becomes $\Pi\underline{x}$, then by the preceding method we obtain

$$\mathcal{E}|(\rho\nu js)\underline{e}_0 n\rangle = (-1)^{j-s_1-s_2}|(\rho-\nu js)\underline{e}_0 n\rangle. \quad (54)$$

If the two particles are identical, we must take into account the symmetry properties of the two-particle states. In this case we have $s_1 = s_2$, and, using the preceeding result, we define two orthonormal combinations, one symmetric and the other antisymmetric, with subscripts S and A respectively:

$$|(\rho\nu js)\underline{e}_0 n\rangle_S = (1/\sqrt{2})[|(\rho\nu js)\underline{e}_0 n_0\rangle + (-1)^{j-2s_1}|(\rho-\nu js)\underline{e}_0 n\rangle], \quad (55a)$$

$$|(\rho\nu js)\underline{e}_0 n\rangle_A = (1/\sqrt{2})[|(\rho\nu js)\underline{e}_0 n\rangle - (-1)^{j-2s_1}|(\rho-\nu js)\underline{e}_0 n\rangle]. \quad (55b)$$

Note that the symmetric (antisymmetric) states are not necessarily associated with bosons (fermions) because internal symmetries must also be considered, for example, isospin for pions.

VI. CONCLUSION

Starting from the one-particle Lorentz basis, we have constructed a "two-particle internal Lorentz basis" of states $|(\rho\nu js)\underline{p}n\rangle$ characterized by:

(i) j the total angular momentum of the two particles in the center of mass, i.e., the spin of the system. It is a Poincaré invariant.

(ii) n is the projection of the spin along the 3-axis.

(iii) s is the vector sum of the two spins s_1 and s_2; it is also a Poincaré invariant.

(iv) $p = P/w$ is the 4-direction of the total 4-momentum of the system of the two particles.

(v,vi) ρ,ν are two internal Lorentz labels, which are invariant under Lorentz transformations but not under the translations. They are the two variables which replace the invariant mass and the helicity of the Poincaré irreducible states. We hope to come back later to the physical significance of these variables.

The states $|\rho\nu js)\underline{p}n_0$ have neither a given invariant mass nor a given 4-momentum, and hence they do not have good transformation properties under the translation group. However, under Lorentz transformations they transform like Poincaré irreducibles states.

These two-particle internal Lorentz bases have been used to derive two-variable expansions of the amplitudes and the observables for two-body scattering with arbitrary spins in terms of the unitary representation matrices of the Lorentz group.[10]

ACKNOWLEDGMENTS

We thank P. Winternitz for various discussions on this subject and related topics.

APPENDIX

In this appendix we give some definitions and results on the unitary representations of the Lorentz group and their Clebsch–Gordan coefficients, which are necessary for the understanding of this article. For a review of the various conventions for the Lorentz basis and for an extended bibliography on this subject the reader is referred to the Appendix of Ref. 10.

Any transformation $\Lambda\in\mathscr{L}^{\uparrow}_{+}$ can be written

$$\Lambda = R(\varphi\theta 0)B(a)R(\alpha\beta\gamma), \tag{A1}$$

where $B(a)$, $a\epsilon[0,\infty]$, is a boost along the 3-axis. The matrix element of the unitary representations specified by $\rho\epsilon[0,\infty)$, $\nu = -\inf(j_1,j_2), \ -\inf(j_1,j_2)+1,\cdots,\inf(j_1,j_2)$, are

$$D^{\rho\nu}_{j_1m_1j_2m_2}(\Lambda) = \sum_m D^{j_1}_{m_1m}(\varphi\theta 0)d^{\rho\nu}_{j_1j_2m}(a)D^{j_2}_{mm_2}(\alpha\beta\gamma), \tag{A2}$$

where the boost matrices $d^{\rho\nu}_{j_1j_2m}(a)$ are defined by

$$\begin{aligned} d^{\rho\nu}_{j_1j_2m}(a) = {} & (-1)^{j_1-j_2}[(2j_1+1)(2j_2+1)]^{1/2} \\ & \times\left[\frac{\Gamma(j_1+i\rho+1)\Gamma(j_2-i\rho+1)}{\Gamma(j_1-i\rho+1)\Gamma(j_2+i\rho+1)}\right]^{1/2} \\ & \times\int_0^1 dt\, d^{j_1}_{\nu m}(2t-1)d^{j_2}_{\nu m}(2t_a-1) \\ & \times[te^{-a}+(1-t)e^{a}]^{i\rho-1} \end{aligned}$$

with $t_a = te^{-a}[te^{-a}+(1-t)e^{a}]^{-1}$ [$d^{j}_{\nu m}$ are O(3) d matrices]. The boost matrices have the following symmetries:

$$d^{\rho\nu}_{j_1j_2\lambda} = d^{-\rho-\nu}_{j_1j_2\lambda}, \tag{A4a}$$

$$d^{\rho\nu*}_{j_1j_2\lambda} = d^{-\rho\nu}_{j_1j_2\lambda}, \tag{A4b}$$

$$d^{\rho\nu}_{j_1j_2\lambda} = d^{\rho\lambda}_{j_1j_2\nu}, \tag{A4c}$$

$$d^{\rho-\nu}_{j_1j_2-\lambda} = d^{\rho\nu}_{j_1j_2\lambda}, \tag{A4d}$$

$$d^{\rho\nu}_{j_1j_2\lambda}(a)^* = d^{\rho\nu}_{j_1j_2\lambda}(-a), \tag{A4e}$$

$$d^{\rho\nu}_{j_1j_2\lambda}(-a) = (-1)^{j_1-j_2}d^{\rho\nu}_{j_1j_2-\lambda}(a). \tag{A4f}$$

The orthonormality and closure relations for the two sets of matrices are [$\mu(\rho,\nu) = \rho^2+\nu^2$]

$$\begin{aligned} \sum_m\int_0^\infty & \sinh^2a\, da\, d^{\rho\nu}_{j_1j_2m}(a)d^{\rho'\nu'}_{j_1j_2m}(a)^* \\ & = \frac{\pi}{2}\frac{(2j_1+1)(2j_2+1)}{\mu(\rho,\nu)}\delta(\rho-\rho')\delta_{\nu\nu'}, \end{aligned} \tag{A5}$$

$$\begin{aligned} \sum_\nu\int_0^\infty & \mu(\rho,\nu)d\rho\, d^{\rho\nu}_{j_1j_2m}(a)d^{\rho\nu}_{j_1j_2m'}(a')^* \\ & = \frac{\pi}{2}\frac{(2j_1+1)(2j_2+2)}{\sinh^2a}\delta(a-a')\delta_{mm'}, \end{aligned} \tag{A6}$$

$$\begin{aligned} \int & \mu(\Lambda)d\Lambda D^{\rho\nu}_{j_1m_1j_2m_2}(\Lambda)D^{\rho'\nu'}_{j_1'm_1'j_2'm_2'}(\Lambda)^* \\ & = \frac{(2\pi)^4}{\mu(\rho,\nu)}\delta(\rho-\rho')\delta_{\nu\nu'}\delta_{j_1j_1'}\delta_{j_2j_2'}\delta_{m_1m_1'}\delta_{m_2m_2'}, \end{aligned} \tag{A7}$$

$$\mu(\Lambda) = \sin\theta\sinh^2a\sin\beta, \quad d\Lambda = d\varphi d\theta da d\alpha d\beta d\gamma,$$

$$\begin{aligned} \sum_{\nu=-\infty}^{+\infty}\int_0^\infty & \mu(\rho,\nu)d\rho\sum_{j=|\nu|}^{\infty}\sum_{m_1m_2}D^{\rho\nu}_{j_1m_1j_2m_2}(\Lambda)D^{\rho\nu}_{j_1m_1j_2m_2}(\Lambda')^* \\ & = \frac{(2\pi)^4}{\mu(\Lambda)}\delta(\Lambda-\Lambda'), \end{aligned} \tag{A8}$$

$$\begin{aligned} \delta(\Lambda-\Lambda') = {} & \delta(\varphi-\varphi')\delta(\theta-\theta')\delta(a-a') \\ & \times\delta(\alpha-\alpha')\delta(\beta-\beta')\delta(\gamma-\gamma'). \end{aligned}$$

The Clebsch-Gordan coefficients of O(3,1) are defined by the tensor product reduction

$$\begin{aligned} & D^{\rho_1\nu_1}_{j_1m_1j_1'm_1'}(\Lambda)D^{\rho_2\nu_2}_{j_2m_2j_2'm_2'}(\Lambda) \\ & = \sum_{\nu=-\infty}^{+\infty}\int_0^\infty\frac{\mu(\rho,\nu)}{(2\pi)^4}d\rho\sum_{j=|\nu|}^{\infty}\sum_m\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1m_1j_2m_2\end{pmatrix}\begin{pmatrix}\rho\nu\\ jm\end{pmatrix} \\ & \quad\times\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1'm_1'j_2'm_2'\end{pmatrix}\begin{pmatrix}\rho\nu\\ j'm'\end{pmatrix}^*D^{\rho\nu}_{jmj'm'}(\Lambda). \end{aligned} \tag{A9}$$

They can be written as the product of an O(3) CG coefficient times a 9j-symbol analyticaly continued

$$\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1m_1j_2m_2\end{pmatrix}\begin{pmatrix}\rho\nu\\ jm\end{pmatrix} = \langle j_1m_1j_2|jm\rangle\begin{bmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1j_2\end{bmatrix}\begin{bmatrix}\rho\nu\\ j\end{bmatrix}, \tag{A10}$$

They satisfy the orthonormality and closure relations

$$\begin{aligned} \sum_{\substack{j_1m_1\\ j_2m_2}}\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1m_1j_2m_2\end{pmatrix}\begin{pmatrix}\rho\nu\\ jm\end{pmatrix}\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1m_1j_2m_2\end{pmatrix}^*\begin{pmatrix}\rho'\nu'\\ j'm'\end{pmatrix}^* \\ = \mu(\rho,\nu)^{-1}\delta(\rho-\rho')\delta_{\nu\nu'}\delta_{jj'}\delta_{mm'}, \end{aligned} \tag{A11}$$

$$\begin{aligned} \sum_{\nu=-\infty}^{+\infty}\int_0^\infty & \mu(\rho,\nu)d\rho\sum_{j=|\nu|}^{\infty}\sum_m\mu(\rho,\nu) \\ & \times\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1m_1j_2m_2\end{pmatrix}\begin{pmatrix}\rho\nu\\ jm\end{pmatrix}\begin{pmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1'm_1'j_2'm_2'\end{pmatrix}^*\begin{pmatrix}\rho\nu\\ jm\end{pmatrix}^* \\ & = \delta_{j_1j_1'}\delta_{m_1m_1'}\delta_{j_2j_2'}\delta_{m_2m_2'}. \end{aligned} \tag{A12}$$

For short we use in the main text the quantities

$$\begin{bmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1j_1'j_2j_2'\end{bmatrix}\begin{bmatrix}\rho\nu\\ jj'\end{bmatrix} = \begin{bmatrix}\rho_1\nu_1\rho_2\nu_2\\ j_1\quad j_2\end{bmatrix}\begin{bmatrix}\rho\nu\\ j\end{bmatrix}\times\begin{bmatrix}\rho_1\nu_2\rho_2\nu_2\\ j_1'\quad j_2'\end{bmatrix}\begin{bmatrix}\rho\nu\\ j'\end{bmatrix}^*, \tag{A13}$$

and Eqs. (A9), (A11), (A12) can be written only by means of these quantities and the O(3) CG coefficients.

[1]E.G. Kalnins, J. Patera, R.T. Sharp, and P. Winternitz, in *Group Theory and Its Applications*, edited by E. Loebl (Academic, New York, 1975), Vol. 3.
[2]M. Daumens and M. Perroud, J. Math. Phys. **20**, 1077 (1979).
[3]E.P. Wigner, Ann. Math. **40**, 149 (1939).
[4]M. Jacob and G.C. Wick, Ann. Phys. (N.Y.) **7**, 404 (1959).
[5]J.S. Lomont and H.E. Moses, J. Math. Phys. **5**, 294, 1438 (1964).
[6]H. Joos, Fortshr. Phys. **10**, 65 (1962).
[7](I) A. Chakrabarti, M. Levy-Nahas, and R. Seneor, J. Math. Phys. **9**, 1274 (1968); (II) A. Chakrabarti, *ibid.* **12**, 1822 (1971).
[8]P. Moussa and R. Stora, in *Angular Analysis of Elementary Particle Reactions, Hercog-Novi* 1966, edited by R. Nikovic´ (Gordon and Breach, New York, 1968).
[9]A.S. Wightman, in *L'invariance dans la mécanique quantique relativiste Les Houches*, 1960 (Hermann, Paris, 1960).
[10]M. Daumens, M. Perroud, and P. Winternitz, "Relativistic Energy Dependent Partial Wave Analysis for Particles with Spin" (to be published), Preprint CEN Saclay, DPhPE78-16.

Reduction of the two-dimensional O(n) nonlinear σ-model

K. Pohlmeyer
Fakultät für Physik der Universität Freiburg, D–7800 Freiburg, Federal Republic of Germany

K.-H. Rehren
Institut für Theoretische Physik der Universität Heidelberg, D–6900 Heidelberg, Federal Republic of Germany

(Received 24 April 1979; accepted for publication 15 June 1979)

We reduce the field equations of the two-dimensional O(n) nonlinear σ-model to relativistic O($n-2$) covariant differential equations involving $n-2$ scalar fields.

I. INTRODUCTION

The classical two-dimensional O(n) nonlinear σ-models define integrable Hamiltonian systems.[1] Taking advantage of conformal invariance, the models corresponding to $n=3$ and $n=4$ can be "reduced" to local relativistic scalar field theories involving O(3) and O(4) invariant combinations of the chiral field vectors and some of their derivatives. The O(3) nonlinear σ-model is reduced to the sine-Gordon theory described by the Lagrangian density[2]

$$\mathscr{L}(x^0,x^1)=\tfrac{1}{2}(\partial_\mu\alpha)(\partial^\mu\alpha)+\cos\alpha-1. \tag{1.1}$$

The O(4) nonlinear σ-model is reduced to a local relativistic theory involving two scalar fields α and β. Its dynamics is described by the Lagrangian density

$$\mathscr{L}(x^0,x^1)=\tfrac{1}{2}(\partial_\mu\alpha)(\partial^\mu\alpha)+\tfrac{1}{2}(\partial_\mu\beta)(\partial^\mu\beta)\tan^2(\alpha/2)+\cos\alpha-1. \tag{1.2}$$

This theory is a generalization of the sine-Gordon theory, where β is identically zero. If we combine α and β into the two-component iso-vector

$$\psi=\sin(\alpha/2)\begin{pmatrix}\cos(\beta/2)\\ \sin(\beta/2)\end{pmatrix},$$

it becomes identical with Getmanov's "New Lorentz-invariant system"[3,4]

$$\mathscr{L}(x^0,x^1)=\frac{1}{2}\frac{(\partial_\mu\psi^a)(\partial^\mu\psi^a)}{1-\psi^a\psi^a}-\frac{1}{2}\psi^a\psi^a. \tag{1.3}$$

The conservation laws and the inverse scattering equations for this "complex sine-Gordon theory" were derived in Ref. 1. Nontopological soliton, multisoliton and breather solutions were obtained in Refs. 3, 5. The transformation to action-angle variables was worked out in Ref. 5.

As can be verified by crossdifferentiation, the Bäcklund transformation mapping solutions ψ of the complex sine-Gordon equation

$$\partial_\mu\partial^\mu\psi+\frac{2(\psi^b\partial_\mu\psi^b)\partial^\mu\psi-(\partial_\mu\psi^b\partial^\mu\psi^b)\psi}{1-\psi^b\psi^b}+(1-\psi^b\psi^b)\psi=0$$

into solutions ψ' of this same equation is

$$R\frac{(\partial_0+\partial_1)\psi'}{\sqrt{1-\psi'^b\psi'^b}}+R^{-1}\frac{(\partial_0+\partial_1)\psi}{\sqrt{1-\psi^b\psi^b}}=\gamma^{-1}\{R^{-1}\psi'\sqrt{1-\psi^b\psi^b}-R\psi\sqrt{1-\psi'^b\psi'^b}\},$$

$$R^{-1}\frac{(\partial_0-\partial_1)\psi'}{\sqrt{1-\psi'^b\psi'^b}}-R\frac{(\partial_0-\partial_1)\psi}{\sqrt{1-\psi^b\psi^b}}=-\gamma\{R\psi'\sqrt{1-\psi^b\psi^b}+R^{-1}\psi\sqrt{1-\psi'^b\psi'^b}\}, \tag{1.4}$$

with γ a real constant parameter different from zero and

$$R=\begin{pmatrix}\cos\omega & -\sin\omega\\ \sin\omega & \cos\omega\end{pmatrix},$$

$$2\omega=\arcsin\left(\frac{\epsilon^{ab}\psi^a\psi'^b}{\sqrt{1-\psi^b\psi^b}\sqrt{1-\psi'^b\psi'^b}}\right).$$

In the present note we shall search for the relativistic differential equations to which the equations of motion of the two-dimensional O(n) nonlinear σ-models can be reduced for higher values of n (c.f., Refs. 6, 7).

II. HIGHER GENERALIZATIONS OF THE SINE-GORDON EQUATION

The classical O(n) nonlinear σ-model describes the motion of a string of n-dimensional classical spins $q^a(x^0,x^1)$, $a=1,\ldots,n$, of unit length: $q^bq^b=1$.[8] The Lagrangian density is

$$\mathscr{L}(x^0,x^1)=\tfrac{1}{2}\{\partial_\mu q^a\partial^\mu q^a+\varkappa(q^aq^a-1)\}, \tag{2.1}$$

where $\varkappa=\varkappa(x^0,x^1)$ is a Lagrangian multiplier. The equations of motion are

$$\partial_\mu\partial^\mu q+(\partial_\mu q^b\partial^\mu q^b)q=0,\quad q^bq^b=1$$
$$[\varkappa(x^0,x^1)=-(\partial_\mu q^b\partial^\mu q^b)].$$

They are invariant under general conformal transformations, space and time reflections, and under the group O(n) of internal rotations and reflections.

We break the conformal invariance by requiring

$$\partial_0q^b\partial_0q^b+\partial_1q^b\partial_1q^b=1,\quad \partial_0q^b\partial_1q^b=0. \tag{2.2}$$

It is advantageous to use light-cone coordinates

$$\xi=(x^0+x^1)/2,\quad \eta=(x^0-x^1)/2,$$

in which the equations of motion and the normalization requirements read

[a]Present address: Fakultät für Physik der Universität Freiburg, D-7800 Freiburg.

 0022-2488/79/122628-05$01.00

$$q_{\xi\eta} + (q^b_\xi q^b_\eta)q = 0, \quad q^b q^b = 1;$$
$$q^b_\xi q^b_\xi = 1 = q^b_\eta q^b_\eta. \tag{2.3}$$

The subscripts ξ and η denote differentiation with respect to ξ and η.

In Ref. 1, in our quest for inverse scattering equations, we started from the Bäcklund transformation for the chiral fields q. In the course of the derivation we obtained two systems of Riccati Eqs. (VII. 11.1) and (VII.11.2) the compatibility of which requires the following relations to hold

$$\alpha_{\xi\eta} + \sin\alpha(s^{(+)}s^{(-)})_{11} = 0,$$
$$(\tan\alpha s^{(+)}_{1j})_\eta + \alpha_\xi s^{(-)}_{1j} + \tan\alpha(s^{(+)}s^{(-)})_{1j} = 0, \quad j = 2,3,\ldots,n-2,$$
$$s^{(+)}_\eta - s^{(-)}_\xi + [s^{(+)}, s^{(-)}] = 0.$$

Here

$$\alpha = \arccos(q^b_\xi q^b_\eta),$$
$$s^{(\pm)} = -s^{(\pm)T[9]}, s^{(\pm)}_{ij} = (b^{\ a}_{i\ \xi} b^{\ a}_{j\ \eta}), \quad i,j = 1,2,\ldots,n-2,$$

with q, q_ξ, $b_1 = (q_\eta - \cos\alpha q_\xi)/\sin\alpha$, b_k; $k = 2,\ldots,n-2$, an orthonormal basis in $\mathbb{R}^n$.

These relations form the starting point of the present investigation. The last equation can immediately be solved:

$$s^{(\pm)}_{ij} = \sum_{b=1}^{n-2} (f^b_{i\xi} f^b_{j\eta}) \tag{2.4}$$

with $f_1 = f, f_2, \ldots, f_{n-2}$ forming an orthonormal basis in $\mathbb{R}^{n-2}$. Now the first two equations read

$$f_{\xi\eta} + \cot\alpha\alpha_\xi f_\eta + (\cos\alpha \sin\alpha)^{-1}\alpha_\eta f_\xi + \sum_{b=1}^{n-2} (f^b_\xi f^b_\eta) f = 0, \tag{2.5a}$$
$$\sum_{b=1}^{n-2} f^b f^b = 1,$$
$$\alpha_{\xi\eta} + \sin\alpha - \tan\alpha \sum_{b=1}^{n-2} (f^b_\xi f^b_\eta) = 0. \tag{2.5b}$$

By setting

$$\sin\alpha f = \varphi$$

they can be combined into a single equation

$$\varphi_{\xi\eta} + \frac{(\varphi\cdot\varphi_\eta)\varphi_\xi}{1 - \|\varphi\|^2} + \sqrt{1 - \|\varphi\|^2}\,\varphi = 0 \tag{2.6}$$

or

$$\partial_\mu\partial^\mu\varphi + \frac{(\varphi\cdot\partial_\mu\varphi)\partial^\mu\varphi + \epsilon^{\mu\nu}(\varphi\cdot\partial_\mu\varphi)\partial_\nu\varphi}{1 - \|\varphi\|^2} + \sqrt{1 - \|\varphi\|^2}\,\varphi = 0. \tag{2.6'}$$

Here the dot denotes the Euclidean scalar product in the space $\mathbb{R}^{n-2}$ and vertical twofold bars stand for the corresponding Euclidean norm.

This equation possesses a one-parameter family of Bäcklund transformations, the transcription of the Bäcklund transformations for the chiral field vectors q to those for φ, and infinitely many local covariant conservation laws, e.g.,

$$\left\{\frac{1}{2}\left|\left|\frac{\varphi_\xi}{\sqrt{1 - \|\varphi\|^2}}\right|\right|^2\right\}_\eta = \left\{\sqrt{1 - \|\varphi\|^2}\right\}_\xi,$$
$$\left\{\tfrac{1}{2}\|\varphi_\eta\|^2 + \frac{1}{2}\frac{(\varphi\cdot\varphi_\eta)^2}{1 - \|\varphi\|^2}\right\}_\xi = \left\{\sqrt{1 - \|\varphi\|^2}\right\}_\eta, \tag{2.7}$$
$$\left\{\frac{1}{2}\left|\left|\left(\frac{\varphi_\xi}{\sqrt{1 - \|\varphi\|^2}}\right)_\xi\right|\right|^2 - \frac{1}{8}\left|\left|\frac{\varphi_\xi}{\sqrt{1 - \|\varphi\|^2}}\right|\right|^4\right\}_\eta = -\left\{\frac{\sqrt{1 - \|\varphi\|^2}}{2}\left|\left|\frac{\varphi_\xi}{\sqrt{1 - \|\varphi\|^2}}\right|\right|^2\right\}_\xi.$$

Had we started from

$$q, q_\eta, b_1 = \frac{q_\xi - \cos\alpha q_\eta}{\sin\alpha}, \quad b_k; k = 2,\ldots,n-2$$

as the orthonormal basis in $\mathbb{R}^n$, we would have obtained in an analogous manner the equation

$$\chi_{\xi\eta} + \frac{(\chi\cdot\chi_\xi)\chi_\eta}{1 - \|\chi\|^2} + \sqrt{1 - \|\chi\|^2}\,\chi = 0 \tag{2.8}$$

or

$$\partial_\mu\partial^\mu\chi + \frac{(\chi\cdot\partial_\mu\chi)\partial^\mu\chi + \epsilon^{\mu\nu}(\chi\cdot\partial_\nu\chi)\partial_\mu\chi}{1 - \|\chi\|^2} + \sqrt{1 - \|\chi\|^2}\,\chi = 0. \tag{2.8'}$$

Though each of the two Eqs. (2.6) and (2.8) possesses infinitely many local covariant conservation laws, none of them can be considered a direct generalization of the real and complex sine-Gordon equation. We shall arrive at such a generalization (for the case $n = 6$) by studying the orthogonal transformation $\mathscr{R}$ mapping the solutions φ of Eq. (2.6) into the solutions χ of Eq. (2.8):

$$\chi = \mathscr{R}\varphi,$$
$$\mathscr{R}_\xi = \frac{-1}{\sqrt{1 - \|\varphi\|^2}\,[1 - \sqrt{1 - \|\varphi\|^2}\,]}\,\mathscr{R}\varphi^a\varphi^{\ b}_\xi I^{\ ab},$$
$$\mathscr{R}_\eta = \frac{-1}{1 - \sqrt{1 - \|\varphi\|^2}}\,\mathscr{R}\varphi^a\varphi^{\ b}_\eta I^{\ ab}, \tag{2.9}$$

where $I^{\ ba} = -I^{\ ab}$ $(a,b = 1,\ldots,n-2)$ denote the infinitesimal generators of the group $O(n-2)$ for rotations in the (a,b) planes. For later convenience we shall work with the covering group. Let Γ^a, $a = 1,\ldots,n-2$, stand for the lowest-dimensional matrix representation of the basis elements of the Clifford algebra[10]

$$\Gamma^a\Gamma^b + \Gamma^b\Gamma^a = 2\delta^{ab}$$

and let the symbol $[\,,\,]$ denote the commutator. The Lie algebra with basis $J^{\ ab} = \frac{1}{4}[\Gamma^a, \Gamma^b]$ is a representation of the Lie algebra of the group $O(n-2)$. The corresponding representatives U of the space-time dependent rotations $\mathscr{R}^T$ satisfy the following equations

$$U_\xi = \frac{1 + \cos\alpha}{\cos\alpha} \sum_{a,b=1}^{n-2} f^a f^b_\xi J^{\ ab} U$$
$$U_\eta = (1 + \cos\alpha) \sum_{a,b=1}^{n-2} f^a f^b_\eta J^{ab} U \tag{2.10}$$
$$UU^+ = U^+ U = \mathbf{1}, \quad \det U = 1.$$

Consistency requires the representatives U to satisfy

$$U_{\xi\eta} + \frac{\alpha_\xi U_\eta + \alpha_\eta U_\xi}{\sin\alpha} + \tfrac{1}{2}\tan^2\frac{\alpha}{2}\,[U_\eta U_\xi^+ - U_\xi U_\eta^+]U + \tfrac{1}{2}[U_\eta U_\xi^+ + U_\xi U_\eta^+]U = 0,$$

$$U^+U = UU^+ = \mathbf{1}, \quad \det U = 1.$$

Equation (2.5b) now reads

$$\alpha_{\xi\eta} + \sin\alpha = \frac{2\tan^2(\alpha/2)}{\sin\alpha}\{U_\eta U_\xi^+ + U_\xi U_\eta^+\}.$$

If we set

$$\sin\frac{\alpha}{2}\,U = V,$$

we obtain

$$V_{\xi\eta} + [1 - VV^+]^{-1}V_\xi V^+ V_\eta + [1 - VV^+]V = 0, \tag{2.11}$$

$$VV^+ = V^+V = \text{multiple of the unit matrix},$$
$$\det V = \text{real}.$$

Independently of its origin, this system possesses an infinite set of local covariant conservation laws, e.g.,

$$\mathrm{Tr}\{\tfrac{1}{2}[1 - VV^+]^{-1}V_\xi V_\xi^+\}_\eta = -\mathrm{Tr}\{\tfrac{1}{2}VV^+\}_\xi \quad (\xi\longleftrightarrow\eta,\ V\longleftrightarrow V^+) \tag{2.12}$$

$$\mathrm{Tr}\{\tfrac{1}{2}[1 - VV^+]^{-1}V_{\xi\xi}V_{\xi\xi}^+ + \tfrac{1}{2}[1 - VV^+]^{-2}V_\xi V_\xi^+ \times [(VV^+)_{\xi\xi} - 4V_\xi V_\xi^+] + \tfrac{1}{2}[1 - VV^+]^{-3} \times (V_\xi V_\xi^+)^2\}_\eta = \mathrm{Tr}\{-V_\xi V_\xi^+(1 - \tfrac{1}{2}[1 - VV^+]^{-1})\}_\xi, \quad (\xi\longleftrightarrow\eta,\ V\longleftrightarrow V^+).$$

The system is likely to be integrable. It contains the solutions of (2.10) as special cases subject to constraints, e.g., for the case $n = 5$ the constraints are

$$[[V_\xi V^+, V_\eta V^+], [V_{\xi\xi}V^+, V_\eta V^+]] = 0,$$
$$[[V_\xi V^+, V_\eta V^+], [V_{\eta\eta}V^+, V_\xi V^+]] = 0.$$

The constraints are simple enough to be resolved only in two cases: $n - 2 = 2$, the case discussed in Refs. 1, 3, 5, and $n - 2 = 4$ [O(4) factorizes!]. In the following, we shall concentrate on the latter case.

III. THE REDUCED EQUATION FOR THE O(6) NONLINEAR σ-MODEL

In this section we shall derive a recursion formula for the conserved current densities valid for all $n \geqslant 3$, and calculate explicitly the first three continuity equations for $n = 6$. Moreover, a general formula for the N-soliton scattering solution is derived. For a special three-soliton configuration it is written in a form which shows the space–time dependence of the field vector most transparently.

In the case under consideration, Eq. (2.10) splits into two sets of equations, each involving an SU(2) matrix. We only need to consider one of them. Parametrizing the SU(2) matrix by a four-dimensional unit vector n, we arrive at

$$n_\xi = -\frac{1 + \cos\alpha}{2\cos\alpha}\{(f_\xi\cdot n)f - (f\cdot n)f_\xi + [n,f,f_\xi]\},$$
$$n_\eta = -\frac{1 + \cos\alpha}{2}\{(f_\eta\cdot n)f - (f\cdot n)f_\eta + [n,f,f_\eta]\}, \tag{3.1}$$

where $[A,B,C]_i = \epsilon_{ijkl}A_jB_kC_l$ denotes the vector product in $\mathbf{R}^4$. Writing $\psi = \sin(\alpha/2)\cdot n$, the compatibility condition for n and the evolution equation (2.5b) for α are cast into the single equation

$$\psi_{\xi\eta} + \frac{(\psi\cdot\psi_\xi)\psi_\eta + (\psi\cdot\psi_\eta)\psi_\xi - (\psi_\xi\cdot\psi_\eta)\psi - [\psi,\psi_\xi,\psi_\eta]}{1 - \|\psi\|^2} + (1 - \|\psi\|^2)\psi = 0. \tag{3.2}$$

The conservation laws are found essentially by a method due to Wadati, Sanuki, Konno.[11] The Riccati equations (VII.11.1) and (VII.11.2) of Ref. 1 yield the continuity equation

$$\left(\frac{\varphi_\xi}{\sqrt{1 - \|\varphi\|^2}}\cdot Z\right)_\eta + \gamma\left((\varphi\cdot Z) - 2\sqrt{1 - \|\varphi\|^2}\right)_\xi = 0$$

where $Z^a = -2\sum_{j=1}^{n-2} f_j^a Y_j$. The expansion of Z in powers of the parameter γ around $\gamma = 0$ and $\gamma = \infty$ leads to two series of conservation laws, e.g., around $\gamma = 0$

$$(Z_1\cdot Z_1)_\eta - 2(\sqrt{1 - \|\varphi\|^2})_\xi = 0,$$
$$(Z_1\cdot Z_{m+1})_\eta + (\varphi\cdot Z_m)_\xi = 0, \quad m \geqslant 1, \tag{3.3}$$

where

$$Z_1 = \frac{\varphi_\xi}{\sqrt{1 - \|\varphi\|^2}},$$

$$Z_{m+1} = (Z_m)_\xi + \sum_{k+l=m}[\tfrac{1}{2}(Z_1\cdot Z_k)Z_l - \tfrac{1}{4}(Z_k\cdot Z_l)Z_1], \quad m \geqslant 1$$

[cf Eq. (2.7) above].

For the case $n = 6$ the first three conservation laws in terms of ψ are[12]

$$\left\{\frac{1}{2}\frac{\|\psi_\xi\|^2}{1 - \|\psi\|^2}\right\}_\eta + \left\{\frac{1}{2}\|\psi\|^2\right\}_\xi = 0,$$

$$\left\{\frac{1}{2}\frac{\|\psi_{\xi\xi}\|^2}{1 - \|\psi\|^2} + \frac{1}{(1 - \|\psi\|^2)^2}\|\psi_\xi\|^2(\psi\cdot\psi_{\xi\xi}) - \frac{1}{2}\frac{1 - 2\|\psi\|^2}{(1 - \|\psi\|^2)^3}\|\psi_\xi\|^4\right\}_\eta + \left\{\frac{1}{2}\frac{1 - 2\|\psi\|^2}{1 - \|\psi\|^2}\|\psi_\xi\|^2\right\}_\xi = 0,$$

$$\left\{\frac{1}{2}\frac{\|\psi_{\xi\xi\xi}\|^2}{1 - \|\psi\|^2} + \frac{1}{(1 - \|\psi\|^2)^2}(\psi\cdot[\psi_\xi,\psi_{\xi\xi},\psi_{\xi\xi\xi}]) + \frac{1}{(1 - \|\psi\|^2)^2}(\psi_{\xi\xi}\cdot\psi_{\xi\xi\xi})(\psi\cdot\psi_\xi)\right.$$

$$
\begin{aligned}
&- \frac{1}{(1-\|\psi\|^2)^2}(\psi_\xi\cdot\psi_{\xi\xi\xi})(\psi\cdot\psi_{\xi\xi}) + \frac{1}{2}\frac{1-3\|\psi\|^2}{(1-\|\psi\|^2)^3}(\psi_\xi\cdot\psi_{\xi\xi\xi})\|\psi_\xi\|^2 + \frac{3}{(1-\|\psi\|^2)^2}(\psi\cdot\psi_{\xi\xi\xi})(\psi_\xi\cdot\psi_{\xi\xi}) \\
&+ \frac{4}{(1-\|\psi\|^2)^3}(\psi\cdot\psi_{\xi\xi\xi})\|\psi_\xi\|^2(\psi\cdot\psi_\xi) - \frac{1}{2}\frac{7-8\|\psi\|^2}{(1-\|\psi\|^2)^3}\|\psi_{\xi\xi}\|^2\|\psi_\xi\|^2 \\
&- \frac{3-7\|\psi\|^2}{(1-\|\psi\|^2)^3}(\psi_\xi\cdot\psi_{\xi\xi})^2 - 4\frac{1-3\|\psi\|^2}{(1-\|\psi\|^2)^4}(\psi_\xi\cdot\psi_{\xi\xi})\|\psi_\xi\|^2(\psi\cdot\psi_\xi) - \frac{1}{2}\frac{15-17\|\psi\|^2}{(1-\|\psi\|^2)^4}(\psi\cdot\psi_{\xi\xi})\|\psi_\xi\|^4 \\
&+ \frac{1}{2}\frac{2-12\|\psi\|^2+11\|\psi\|^4}{(1-\|\psi\|^2)^5}\|\psi_\xi\|^6 - \frac{1-5\|\psi\|^2}{(1-\|\psi\|^2)^5}\|\psi_\xi\|^4(\psi\cdot\psi_\xi)^2\Big\}_\eta \\
&+ \Big\{\frac{1}{2}\frac{1-2\|\psi\|^2}{1-\|\psi\|^2}\|\psi_{\xi\xi}\|^2 - \frac{5}{1-\|\psi\|^2}(\psi_\xi\cdot\psi_{\xi\xi})(\psi\cdot\psi_\xi) + \frac{1}{2}\frac{1-3\|\psi\|^2}{(1-\|\psi\|^2)^2}(\psi\cdot\psi_{\xi\xi})\|\psi_\xi\|^2 \\
&- \frac{1}{2}\frac{(2-3\|\psi\|^2)^2}{(1-\|\psi\|^2)^2}\|\psi_\xi\|^4 - \frac{5}{(1-\|\psi\|^2)^2}\|\psi_\xi\|^2(\psi\cdot\psi_\xi)^2\Big\}_\xi = 0. \qquad (3.4)
\end{aligned}
$$

A second series is found by interchanging ($\xi \longleftrightarrow \eta$) and ($[A,B,C] \longleftrightarrow -[A,B,C]$).

Equations (VIII.9) and (VIII.10) of Ref. 1 allow us to apply the inverse scattering method to solve the differential Eq. (3.2). Let us write the linear operators L and B in the form

$$
L(\eta) = i\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\partial_\xi - \frac{i}{2}\begin{pmatrix}0 & \varphi_\xi/\sqrt{1-\|\varphi\|^2}\\ \varphi_\xi^{+}/\sqrt{1-\|\varphi\|^2} & 0\end{pmatrix},
$$

$$
B = \frac{\gamma}{2}\begin{pmatrix}-\sqrt{1-\|\varphi\|^2}\,1 & \varphi\\ \varphi^{+} & \sqrt{1-\|\varphi\|^2}\,1\end{pmatrix}
$$

where

$$
\varphi = -i\sigma^1\varphi^1 - i\sigma^2\varphi^2 - i\sigma^3\varphi^3 + 1\varphi^4 .
$$

In a similar way as was worked out by Takhtadzhyan to calculate the N-soliton-scattering in the sine-Gordon theory,[13] we find for $r(\lambda) = 0$ the "scattering potential"

$$
\frac{\varphi_\xi}{\sqrt{1-\|\varphi\|^2}} = -4\,\mathrm{tr}\{[1 - W_-(\xi,\eta)W_+(\xi,\eta)]^{-1}\partial_\xi W_-(\xi,\eta)\}, \qquad (3.5)
$$

$$
\varphi = -\left(\frac{\varphi_\xi}{\sqrt{1-\|\varphi\|^2}}\right)_\eta .
$$

Here W_+ are GL(2,$\mathbb{C}$)-valued $N\times N$ matrices with entries

$$
W_{-jk}(\xi,\eta) = \frac{1}{\varkappa_j+\varkappa_k}\exp\left((\varkappa_j+\varkappa_k)\xi - \frac{1}{2\varkappa_j}\eta\right)m_j,
$$

$$
W_{+jk}(\xi,\eta) = \frac{-1}{\varkappa_j+\varkappa_k}\exp\left((\varkappa_j+\varkappa_k)\xi - \frac{1}{2\varkappa_j}\eta\right)(m_j)^{+}.
$$

$\varkappa_j$ are N different arbitrary complex numbers with $\mathrm{Re}\varkappa_j > 0$, $\varkappa_{j'} \equiv (\varkappa_j)^*$, and m_j are arbitrary constant GL(2,$\mathbb{C}$) matrices subject to the symmetry relation $m_{j'} = \sigma^2 m_j^* \sigma^2$, which is due to a symmetry of the scattering operators $L(\eta)$. The trace "tr" denotes the sum over the diagonal matrix-valued entries of the $N\times N$ matrix. For a more detailed derivation see Ref. 12. The pairs $(\varkappa_j,\varkappa_{j'})$ correspond to N_B breathers in the asymptotic state of the solution, the $N_S = N - 2N_B$ real $\varkappa_j$ are related to solitons. φ and ψ depend on the vectors m_j in an SO(4)-covariant manner. Only the real parts of m_j survive in Eq. (3.5). Hence there are just $N_S + N_B$ independent vectors available to build the space in which ψ develops. Thus, the simplest solution of Eq. (3.2), exhibiting, however, those features which are characteristic for the case $n = 6$, is the three-soliton scattering solution.

We present this solution for the case where the polarizations of the solitons are mutually perpendicular ($m_i = -i\sigma^i M_i$, $i = 1,2,3$):

$$
\begin{aligned}
\psi_i(\xi,\eta) &= \frac{(M_i/\varkappa_i)E_i(1-\Sigma_j c_{ij}E_j^2 + c_i E_j^2 E_k^2)}{1+\Sigma_j a_j E_j^2 + \Sigma_{jk} a_{jk}E_j^2E_k^2 + a_{123}E_1^2E_2^2E_3^2}, \quad i = 1,2,3, \\
\psi_4(\xi,\eta) &= \frac{(M_1M_2M_3/\varkappa_1\varkappa_2\varkappa_3)E_1E_2E_3(\varkappa_1-\varkappa_2)(\varkappa_2-\varkappa_3)(\varkappa_3-\varkappa_1)/(\varkappa_1+\varkappa_2)(\varkappa_2+\varkappa_3)(\varkappa_2+\varkappa_1)}{1+\Sigma_j a_j E_j^2 + \Sigma_{jk} a_{jk}E_j^2E_k^2 + a_{123}E_1^2E_2^2E_3^2} \qquad (3.6)
\end{aligned}
$$

with the notation

$$
E_i(\xi,\eta) = \exp\left(2\varkappa_i\xi - \frac{1}{2\varkappa_i}\eta\right) = \exp\left[\left(\varkappa_i + \frac{1}{4\varkappa_i}\right)x^1 + \left(\varkappa_i - \frac{1}{4\varkappa_i}\right)x^0\right],
$$

$$c_{ij} = = \frac{M_j^2}{4\kappa_j^2}\frac{(\kappa_i-\kappa_j)^2}{(\kappa_i+\kappa_j)^2},$$

$$c_i = \frac{M_j^2 M_k^2}{16\kappa_j^2\kappa_k^2}\frac{(\kappa_i-\kappa_j)^2(\kappa_i-\kappa_k)^2(\kappa_j-\kappa_k)^4}{(\kappa_i+\kappa_j)^2(\kappa_i+\kappa_k)^2(\kappa_j+\kappa_k)^4}, j\neq i\neq k;$$

$$a_i = \frac{M_i^2}{4\kappa_i^2},$$

$$a_{ij} = \frac{M_i^2 M_j^2}{16\kappa_i^2\kappa_j^2}\frac{(\kappa_i-\kappa_j)^4}{(\kappa_i+\kappa_j)^4},$$

$$a_{123} = \frac{M_1^2 M_2^2 M_3^2}{64\kappa_1^2\kappa_2^2\kappa_3^2}\frac{(\kappa_1-\kappa_2)^4(\kappa_2-\kappa_3)^4(\kappa_3-\kappa_1)^4}{(\kappa_1+\kappa_2)^4(\kappa_2+\kappa_3)^4(\kappa_3+\kappa_1)^4}.$$

We observe that—at least in this example—we can write

$$\psi = 2\,\mathrm{tr}\{[1 - W_- W_+]^{-1}\partial_\xi \partial_\eta W_-\}.$$

IV. CONCLUSIONS

The field equations of the two-dimensional nonlinear O(n) σ-model can be reduced to either one of two systems of relativistic differential equations involving $(n-2)$ scalar fields in an O($n-2$) covariant manner. Both systems possess a denumerably infinite set of local covariant conservation laws.

The representative of the space–time dependent rotation mediating between the two reduced field vectors itself satisfies a differential equation invariant under the restricted Poincaré group. For $n-2=4$ a recursion formula for its local conservation laws is derived. A formula for explicitly calculating multisoliton scattering solutions is given.

[1] K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).

[2] Notation: Greek indices μ,ν run from 0 to 1, Latin ones a,b from 1 to n if not stated otherwise. The metric is $g_{00} = -g_{11} = 1$. $\epsilon^{\mu\nu}$ is the antisymmetric tensor $\epsilon^{01} = -\epsilon^{10} = 1$. The summation convention is implied.

[3] B.S. Getmanov, Pis'ma Zh. Eksp. Teor. Fiz. **25**, 132 (1977) [Sov. Phys. JETP Lett. **25**, 119 (1977)].

[4] By a rescaling of the coordinates and the fields, a two-parameter theory is obtained corresponding to the Lagrangian density

$$\mathscr{L}(x^0,x^1) = \frac{1}{2}\frac{(\partial_\mu\psi^a)(\partial^\mu\psi^a)}{1-\lambda^2\psi^a\psi^a} - \frac{m^2}{2}\psi^a\psi^a.$$

[5] F. Lund, Phys. Rev. Lett. **38**, 1175 (1977).

[6] A.S. Budagov and L.A. Takhtadzhyan, Dokl. Akad. Nauk SSSR **235**, 805 (1977) (in Russian).

[7] V.E. Zakharov and A.V. Mikhailov, Zh. Eksp. Teor. Fiz. **74**, 1953 (1978) (in Russian).

[8] M. Lüscher and K. Pohlmeyer, Nucl. Phys. B **137**, 46 (1978).

[9] The symbols T, *, and $+$, the latter ones to be employed later, denote transposition, complex and Hermitian conjugation respectively.

[10] H. Boerner, *Darstellungen von Gruppen* (Springer-Verlag, Berlin, Göttingen, Heidelberg, 1955).

[11] M. Wadati, H. Sanuki, and K. Konno, Prog. Theor. Phys. **53**, 419, §5 (1975).

[12] K.-H. Rehren, Heidelberg Preprint HD-THEP 78-22 (1978). Content: derivation of the local conservation laws, formulation of the inverse scattering method, and detailed construction of soliton scattering solutions for the reduced equations of the O(6) nonlinear σ-model.

[13] L.A. Takhtadzhyan, Zh. Eksp. Teor. Fiz. **66**, 476 (1974) [Sov. Phys. JETP **39**, 228 (1974)].

Albedo problem of radiative transfer in inhomogeneous isotropically scattering atmospheres: Application of the maximum variational principle

Madhoo Kanal

Department of Physics, Clark University, Worcester, Massachusetts 01610
(Received 6 December 1978)

An infinite medium Green's function, constructed from a maximum variational principle, is used to solve the albedo problem of radiative transfer in an optically inhomogeneous medium. We assume that the medium is isotropically scattering; however, the albedo for single scattering is kept as an arbitrary continuous function of the optical depth. For a finite medium (such as the slab problem in neutron transport theory) we recommend that use of some experimental measurements be made to circumvent the mathematical difficulty of solving coupled singular integral equations. For a semi-infinite medium we obtain analytic solution for the albedo problem in the framework of the trial Green's function obtained from the maximum variational principle. For the basis functions we make heavy use of Case's eigenfunctions with a slight generalization to account for the inhomogeneous character of the medium. We also make heavy use of our previous work (c f. Kanal and Moses, a reference in this paper) which should be consulted in conjunction with the present work.

1. INTRODUCTION

Since the times of Rayleigh, Schuster, Milne, Eddington, and subsequently Chandrasekhar, much attention has been given to the transfer of radiation to the special cases where the atmospheric properties are so specialized that the equation of radiative transfer lends itself to reasonable tractable analytical solutions. For instance, such is the case when the medium is optically homogeneous so that, from the mathematical point of view, the equation of transfer becomes separable in the angular and optical coordinates. In this regard, *inter alia*, one has Case's[1] powerful normal mode expansion technique, Chandrasekhar's[2] discrete ordinate method, or Ambarzumian's[3] method of invariant imbedding. For the classical works on radiative transfer, one should refer to the excellent collection of collected papers given in the book edited by Menzel.[4]

When the separability criterion cannot be met, and generally it is not, there is unfortunately no direct way of adopting these powerful techniques in solving the boundary-value problems. One example of this case, where the equation of transfer is not separable, is when a plane parallel atmosphere is vertically inhomogeneous so that the albedo for single scattering bears some arbitrary dependence on the optical depth. In particular, for the terrestrial atmosphere, the vertical profile of the neutral species density distribution is such that there is no one, single well-defined function which characterizes the albedo for single scattering. Inherent in the boundary value problems is also the difficulty of solving the appropriate singular integral equations for the expansion coefficients. For instance, even for the seemingly simple case of the albedo problem for a finite homogeneous isotropically scattering atmosphere bounded by two optical planes, the best one can do is to solve the integral equations iteratively. This is the finite slab problem encountered in neutron transport as well and is discussed in Ref. 1. Only for the semi-infinite case is the exact solution known. In view of these difficulties, we present here a pratical way of dealing with the albedo problem for an optically finite atmosphere. For the present we restrict our treatment to the case of isotropic scattering, but allow the albedo for single scattering to bear an arbitrary dependence on the optical depth. Our procedure is as follows: In a paper by Kanal and Moses,[5] the boundary value problem was set up in a manner which places maximum burden of difficulty on the construction of a single infinite medium Green's function for a rather large class of problems.[1,6] Thus, a greater flexibility is achieved with respect to boundary conditions. As in the classical problems of electrodynamics, we find that, in the integral representation of the specific intensity, only the boundary values of the radiation field occur in association with the Green's function [cf. Eq. (10) in Ref. 5]. This fact is, of course, well known. To solve for the radiation field everywhere, we propose the obvious, that the boundary value of the radiation be provided by an actual measurement at one of the optical planes bounding the atmosphere. For the terrestrial atmosphere, for instance, that would consist of ground-based measurement of the specific intensity for all angles for the chosen frequency of radiation. Then knowing the incident flux at the top of the atmosphere, one can solve for the reflected intensity and in consequence determine the state of the radiation field everywhere in the atmosphere. In the case when the suitable experiments are set up in such a manner that one has simultaneous information on radiation at both optical planes, then the problem of determining the radiation field everywhere reduces to a mere angular integration of the infinite medium Green's function appropriately weighted by the boundary values. In any event we are always faced with the problem of

 0022-2488/79/122633-06$01.00

constructing the Green's function. In Ref. 5 we have used a maximum variational principle to circumvent, at least partially, this difficulty. Our present treatment of the finite albedo problem relies entirely on the results obtained in that reference. For that reason we recommend to the reader that Ref. 5 be consulted before proceeding with this paper any further beyond this section.

2. INTEGRAL REPRESENTATION OF THE SPECIFIC INTENSITY

The central problem we are addressing is the albedo problem, which requires obtaining the radiation field everywhere in a source free optically finite medium, if a parallel beam is incident on one optical plane and the character of radiation entering the medium at the other plane is prescribed. This is strictly the albedo problem. We shall, however, set up a more general boundary value problem so that the procedure given here can be applied to a larger class of problems; including the slab problem of neutron transport. To focus our attention, we shall use the terrestrial model and set up the coordinate system accordingly.

In the usual real coordinate, if z represents the altitude measured from the ground up and μ the cosine of the angle with respect to the positive z axis, then the equation of radiative transfer for an azimuthally symmetric plane parallel atmosphere may be written as

$$\mu \frac{\partial}{\partial z} I(z,\mu) + \lambda_t^{-1} I(z,\mu) = \frac{\lambda_e^{-1}}{2} \int_{-1}^{1} d\mu' I(z,\mu'), \tag{1}$$

where $I(z,\mu)$ is the specific intensity of radiation in Rayleighs, λ_t is the mean free path of photons corresponding to all possible photon-matter interactions, and λ_e is the mean free paths of photons corresponding to the elastic scattering. The inverse mean free paths are also referred to as the total absorption and elastic scattering coefficients, respectively, and are defined as the product of the number density and the cross section summed over all the species and their quantum states. Thus, explicitly

$$\lambda_e^{-1} = \sum_I n_I(z) Q_I^{(e)}(\nu), \tag{2}$$

$$\lambda_t^{-1} = \sum_I n_I(z) Q_I^{(t)}(\nu), \tag{3}$$

where $n_I(z)$ is number density as a function of altitude z, $Q_I^{(e)}(\nu)$ is the frequency (ν) dependent elastic scattering cross section, and $Q_I^{(t)}(\nu)$ is the total absorption cross section. The subscript I in all quantities refers to the I th species. Since we are dealing with incoherent radiative transfer, the frequency ν appears as a parameter, and we omit the explicit use of that symbol. We define the optical depth, x, in the usual manner,

$$x = \int_z^\infty dz' \lambda_t^{-1} = \int_z^\infty dz' \sum_I n_I(z') Q_I^{(t)} \tag{4}$$

and the albedo for single scattering $\omega(z)$ as

$$\omega(z) = \lambda_e^{-1}/\lambda_t^{-1} = \sum_I n_I(z) Q_I^{(e)} \Big/ \sum_I n_I(z) Q_I^{(t)}. \tag{5}$$

From definitions of x and $\omega(z)$ we see that in general ω will be a function of x, which by definition characterizes an inhomogeneous atmosphere. If the atmosphere consists of a single species, then ω is independent of the density and reduces to a mere ratio of scattering and total cross sections. In other words, the fraction of photons, which are elastically scattered on a single encounter is independent of the density distribution of the single species. This is also true for a multispecies atmosphere provided for any given frequency of radiation the medium is conservative; i.e., there is no true absorption and the only process of removal of photons from the main beam is by pure elastic scattering process. In that case ω is unity (conservative). For the Venus atmosphere, ω is very close to unity ($\omega \simeq 0.99$) in the visible as is evident from its extreme brightness.

We now set up the general boundary-value problem for a finite atmosphere under the assumption that we know the boundary conditions at the two surfaces optically bounding the atmosphere. In terms of the optical scale defined by Eq. (4) and $\omega(x)$ defined by Eq. (5), the equation of transfer (1) takes the standard form given by

$$\mu \frac{\partial I}{\partial x}(x,\mu) + I(x,\mu) = \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' I(x,\mu') + Q(x,\mu), \tag{6}$$

where we have added a source term $Q(x,\mu)$, which will always be assumed known, and also changed the sign of μ so that in the optical scale μ is positive for the radiation entering the atmosphere from the top, i.e., $z = \infty$. This is purely for the reason of adhering to the notation used in Ref. 5. In the same reference we have shown that if one considers a time reversed adjoint equation

$$\begin{aligned} -\mu \frac{\partial}{\partial x} \tilde{G}(x, -\mu \to x_0, -\mu_0) + \tilde{G} &= \frac{\omega(x)}{2} \int_{-1}^{1} d\mu' \tilde{G}(x, -\mu' \to x_0, -\mu_0) \\ &+ \frac{[\omega(x)]^{1/2}}{2} \delta(x - x_0)\delta(\mu - \mu_0), \end{aligned} \tag{7}$$

which is the Green's function, then for a finite atmosphere optically bounded between $0 \leqslant x \leqslant x_b$ the appropriate combination of Eqs. (6) and (7) leads to an integral representation for $I(x,\mu)$ given by

$$\begin{aligned} I(x,\mu)\Theta(x)\Theta(x_b - x) &= \frac{2}{[\omega(x)]^{1/2}} \int_{-1}^{1} d\mu' \mu' [I(0,\mu')\tilde{G}(0, -\mu' \to x, -\mu) \\ &- I(x_b,\mu')\tilde{G}(x_b, -\mu' \to x, -\mu)] + \frac{2}{[\omega(x)]^{1/2}} \\ &\times \int_0^{x_b} dx' \int_{-1}^{1} d\mu' \tilde{G}(x', -\mu' \to x, -\mu) Q(x',\mu'), \end{aligned} \tag{8}$$

where $\Theta(x)$ is the Heaviside step function. We also showed that $\tilde{G}$ satisfies a reciprocity relation

$$[\omega(x)]^{1/2} \tilde{G}(x, -\mu \to x_0, -\mu_0) = [\omega(x_0)]^{1/2} G(x_0,\mu_0 \to x,\mu), \tag{9}$$

where $G(x_0,\mu_0\to x,\mu)$ satisfies the following equation:

$$\mu\frac{\partial}{\partial x}G(x_0,\mu_0\to x,\mu)+G = \frac{\omega(x)}{2}\int_{-1}^{1}d\mu'\,G(x_0,\mu_0\to x,\mu') + \frac{[\omega(x)]^{1/2}}{2}\delta(x-x_0)\delta(\mu-\mu_0). \tag{10}$$

It is worthwhile to point out that, in view of the lack of translation invariance of $\tilde{G}$ and G, it is important to keep in mind the order of arguments of both Green's functions $\tilde{G}$ and G. Thus in $\tilde{G}(G)$ the right (left) part is the location of the plane source. In Eq. (8) we may now replace the adjoint Green's functions $\tilde{G}$ by G with the use of the reciprocity relations (9). This leads to the integral representation for $I(x,\mu)$ of the form

$$I(x,\mu)\Theta(x)\Theta(x_b-x) = 2\int_{-1}^{1}d\mu'\mu'\left[\frac{1}{[\omega(0)]^{1/2}}I(0,\mu')G(x,\mu\to x_b,\mu') - \frac{1}{[\omega(x_b)]^{1/2}}I(x_b,\mu')G(x,\mu\to x_b,\mu')\right] + 2\int_0^{x_b}dx'\int_{-1}^{1}d\mu'\frac{Q(x',\mu')}{[\omega(x')]^{1/2}}G(x,\mu\to x',\mu'). \tag{11}$$

From the integral identity (11) it is clear that the state of the radiation field is completely determined when its boundary values $I(0,\mu)$, $I(x_b,\mu)$, and any internal sources $Q(x,\mu)$ are known for all angles. This is provided we can construct the Green's functions G from the solution of Eq. (10). If one assumes for the moment that G is known, then from the integral identity (11) we obtain the integral equations for the reflected intensity [i.e., $I(0,\mu)$, $\mu<0$] and the transmitted intensity [i.e., $I(x_b,\mu)$, $\mu>0$] under the boundary conditions that we know the incident intensity [i.e., $I(0,\mu)$, $\mu>0$] and the intensity entering from the bottom of the atmosphere [i.e., $I(x_b,\mu)$, $\mu<0$]. Thus, in the limit that x approaches $x=0$ and $x=x_b$, we obtain

$$I(0,\mu) = 2\int_{-1}^{1}d\mu'\,\mu\left[\frac{I(0,\mu')}{[\omega(0)]^{1/2}}G^{+}(0,\mu\to 0,\mu') - \frac{I(x_b,\mu')}{[\omega(x_b)]^{1/2}}G(0,\mu\to x_b,\mu')\right] + 2\int_0^{x_b}dx'\int_{-1}^{1}d\mu'\frac{Q(x',\mu')}{[\omega(x')]^{1/2}}G(0,\mu\to x',\mu'), \tag{12}$$

$$I(x_b,\mu) = 2\int_{-1}^{1}d\mu'\left[\frac{I(0,\mu')}{[\omega(0)]^{1/2}}G(x_b,\mu\to 0,\mu') - \frac{I(x_b,\mu')}{[\omega(x_b)]^{1/2}}G^{-}(x_b,\mu\to x_b,\mu')\right] + 2\int_0^{x_b}dx' \times\int_{-1}^{1}d\mu'\,Q(x',\mu')\frac{1}{[\omega(x')]^{1/2}}G(x_b,\mu\to x',\mu'), \tag{13}$$

where $+\,(-)$ on $G^{\pm}$ represents the boundary value of G as $x\to 0+$ ($x\to x_b-0$, i.e, from left of x_b).

Thus in principle one may solve integral equations (12) and (13) under the aforementioned boundary conditions and find $I(0,\mu)$ for $\mu<0$ and $I(x_b,\mu)$ for $\mu>0$ and use those in Eq. (11) to find $I(x,\mu)$ everywhere. Analytically, a rigorous solution for a finite atmosphere has not been obtained. Even for the case when ω is constant, only two limiting cases lend to some useful iterative solutions. One is when the atmosphere is optically very thick $x_b\gg 1$, and the other is when it is optically very thin $x_b\ll 1$. Only for the semi-infinite atmosphere is a rigorous solution known.[1,2] With the additional complication of a variable ω, the situation becomes even more hopeless. We, therefore, propose to combine the analytical method with the experimental measurements and require that the measurements be of sufficient accuracy to test the self-consistency of integral equations (12) and (13). We discuss two cases in detail:

(a) when both $I(0,\mu)$ and $I(x_b,\mu)$ are given by measurement for all angles, and

(b) when either of $I(0,\mu)$ or $I(x_b,\mu)$ is given for all angles.

In both cases we shall require that the incident radiation be a known quantity. As our difficulty then lies in constructing the appropriate Green's function from Eq. (10), we, therefore, discuss the resolution of that problem first as recommended in Ref. 5. In particular, for case (b) we shall require that the planetary reflection condition be given, i.e.,

$I(x_b,\mu)$ = known conditions for $\mu<0$.

3. AN APPROXIMATE GREEN'S FUNCTION

We choose a trial Green's functions which maximizes a certain functional. Thus, following Ref. 5, we choose the trial Green's function G_T [see Eq. (59) in Ref. 5] such that

$$G_T(x_0,\mu_0\to x,\mu) = \alpha(x_0)\frac{\omega(x_0)}{\omega(x)}G_g(x_0,\mu_0\to x,\mu), \tag{15}$$

where G_g is the solution of Eq. (10) with $\omega(x)$ replaced by $\omega(x_0)$ (here x_0 is the source point) and $\alpha(x_0)$ is a scaling parameter determined by the variational principle and is given by [see Eq. (66) in Ref. 5]

$$\alpha(x_0) = [1+W(x_0)/\tilde{F}[\rho_{gS}]]^{-1}, \tag{16}$$

where

$$W(x_0) = \int_{-\infty}^{\infty}dx\left(\frac{\omega(x_0)}{\omega(x)}-1\right)\rho_{gS}^2(x_0,x), \tag{17}$$

and $\rho_{pS}(x_0,x)$ is the weighted density function for the plane source located at x_0 and is defined as [see Eq. (56) in Ref. 5]

$$\rho_{gS}(x_0,x) = \frac{\omega(x_0)}{2}\left(\frac{\exp(-|x-x_0|/\nu_0)}{N_{0+}(x_0)} + \int_0^1\frac{d\nu}{N(\nu,x_0)}\exp(-|x-x_0|/\nu)\right).$$

The quantity $\widetilde{F}[\rho_{gS}]$ in Eq. (16) is the value of the functional [see Eq. (64) in Ref. 5]

$$\widetilde{F}[\rho_{gS}] = \int_{-\infty}^{\infty} dx\, K_g(x,x_0)\,\rho_{gS}(x_0,x), \tag{19}$$

where the kernel $K_g(x,x_0)$ is defined by

$$K_g(x,x_0) = \frac{\omega(x_0)}{2} E_1(|x-x_0|) \tag{20}$$

with $E_1(|x-x_0|)$ representing the exponential integral, i.e.,

$$E_1(x-x_0|) = \int_0^1 \frac{d\mu}{\mu} \exp\left(\frac{-|x-x_0|}{\mu}\right) \tag{21}$$

Finally, the quantity ν_0 in (18) is the zero of the dispersion function

$$1 - \frac{\omega(x_0)}{2}\nu_0(x_0)\cdot \ln\left(\frac{\nu_0(x_0)+1}{\nu_0(x_0)-1}\right) = 0 \tag{22}$$

defining a trajectory as x_0 changes. The quantites $N_{0\pm}(x_0)$, $N(\nu,x_0)$ in (18) are the normalization coefficients of Case's discrete and continuum eigenfunctions defined below. The Green's function G_g in (15) has the representation [see Eq. (48) in Ref. 5]

$$\begin{aligned} &G_g(x_0,\mu_0\to x,\mu) \\ &= \pm \frac{[\omega(x_0)]^{1/2}}{2}\left[\phi_{0\pm}(\mu,x_0)\phi_{0\pm}(\mu_0,x_0)\right. \\ &\quad\times\exp\left(\frac{-|x-x_0|}{\nu_0}\right) + \int_0^1 \frac{d\nu}{N(\pm\nu,x_0)} \\ &\quad\left.\times\phi_{\pm\nu}(\mu,x_0)\phi_{\pm\nu}(\mu_0,x_0)\exp\left(\frac{-|x-x_0|}{\nu}\right)\right], \end{aligned} \tag{23}$$

where $\pm$ ($-$) is for $x > x_0$ ($x_0 < x$) and $\phi_{0\pm}(\mu,x_0)$, $\phi_\nu(\mu,x_0)$ are discrete and continuum Case-eigenfunctions given by

$$\phi_{0\pm}(\mu,x_0) = \frac{\nu_0(x_0)\omega(x_0)}{2[\pm\nu_0(x_0)-\mu]}, \tag{24}$$

$$\phi_\nu(\mu,x_0) = \frac{\nu\omega(x_0)}{2}\,\mathcal{P}\frac{1}{\nu-\mu} + \lambda(\nu,x_0)\delta(\nu-\mu), \tag{25}$$

$$\lambda(\nu,x_0) = 1 - \nu\omega(x_0)\tanh^{-1}\nu, \tag{26}$$

$$N_{0\pm}(x_0) = \frac{\omega(x_0)}{2}\nu_0^3(x_0)\left(\frac{\omega(x_0)}{\nu_0^2(x_0)-1} - \frac{1}{\nu_0^2(x_0)}\right). \tag{27}$$

$$N(\nu,x_0) = \nu[(1-\omega(x_0)\nu\tanh^{-1}\nu)^2 + \omega^2(x_0)\pi^2\nu^2/4], \tag{28}$$

and

$$N_{0-} = -N_{0+}.$$

We notice that in the spectral representation of the Green's function, given by (23), the discrete and continuum eigenfunctions involve the argument x_0. This is a fact of great importance to keep in mind for the reason that the way G_g was constructed from Eq. (10), we replaced $\omega(x)$ by $\omega(x_0)$, which corresponds to saying that G_g is an infinite medium Green's function for that medium whose scattering and absorption properties are copies of the "local" properties where the plane source is located, i.e., at $x = x_0$. In other words, one has a continuum ensemble of "source functions" representing the actual medium. One may ask the question as to the nature of the absolute uniform error (i.e., for all x_0) introduced in such a representation. We have addressed that question from the point of view of the maximal variational principle and given an estimate of that error [see Eq. (68) in Ref. 5], which is

$$\epsilon(x_0) = \{F[\rho_{TS}] - \widetilde{F}[\rho_{gS}]\}/F[\rho_{TS}], \tag{29}$$

where $\widetilde{F}[\rho_{gS}]$ is the functional given by (19) and the functional $F[\rho_{TS}]$ is defined by [see Eq. (63) in Ref. 5]

$$F[\rho_{TS}] = \alpha(2-\alpha)\widetilde{F}[\rho_{gS}] - \alpha^2 W(x_0) \tag{30}$$

Here

$$\begin{aligned} &\rho_{TS}(x_0,x) = [\omega(x)]^{1/2}\rho_T(x_0,x), \\ &\rho_T(x_0,x) = \int_{-1}^{1} d\mu \int_{-1}^{1} d\mu_0\, G_T(x_0,\mu_0\to x,\mu). \end{aligned} \tag{31}$$

Thus, for the trial function G_T chosen here, Eq. (29) should give a fairly accurate estimate of the confidence factor in that choice. A more general and, in fact, an exact bound on the error will be presented in the near future in the context of the solution of the Gel'fand–Leviton equation. We finally wish to point out that in the choice of our notation the argument in $G_g(x_0,\mu_0\to x,\mu)$ is such that the left part (x_0,μ_0) represents the location of the plane source. We follow this convention throughout the present treatment.

4. ALBEDO PROBLEM

Having obtained an approximate Green's function G_T, given by (15) with G_g given by (23), we go back to the integral representation (11) of the specific intensity $I(x,\mu)$ and replace G by G_T and write

$$\begin{aligned} &I_T(x,\mu)\Theta(x)\Theta(x_b-x) \\ &= 2\alpha(x)\left[\frac{\omega(x)}{\omega^{3/2}(0)}\int_{-1}^{1} d\mu'\,\mu' I(0,\mu')G_g(x,\mu\to 0,\mu')\right. \\ &\quad - \frac{\omega(x)}{\omega^{3/2}(x_b)}\int_{-1}^{1} d\mu'\,\mu' I(x_b,\mu')G_g(x,\mu\to x_b,\mu') \\ &\quad \left. + \int_0^{x_b} dx\,\frac{\omega(x)}{\omega^{3/2}(x')}\int_{-1}^{1} d\mu'\,Q(x',\mu')G_g(x,\mu\to x',\mu')\right], \end{aligned} \tag{32}$$

where we have subscripted $I(x,\mu)$ with T to indicate the use of a trial Green's function and used the relation (15) between G_T and G_g to express the right-hand side in terms of G_g. It is convenient to introduce the following quantities.

$$A(x,\mu|0) = \frac{2}{[\omega(x)]^{1/2}}\int_{-1}^{1} d\mu'\,\mu' I(0,\mu')G_g(x,\mu\to 0,\mu'), \tag{33}$$

$$A(x,\mu|x_b) = \frac{2}{[\omega(x)]^{1/2}}\int_{-1}^{1} d\mu'\,\mu' I(x_b,\mu')G_g(x,\mu\to x_b,\mu'), \tag{34}$$

$$B(x,\mu|x') = \frac{2}{[\omega(x)]^{1/2}}\int_{-1}^{1} d\mu'\,Q(x',\mu')G_g(x,\mu\to x',\mu'), \tag{35}$$

$$R(x|0) = \left[\frac{\omega(x)}{\omega(0)}\right]^{3/2}, \tag{36}$$

$$R(x|x_b) = \left[\frac{\omega(x)}{\omega(x_b)}\right]^{3/2}, \tag{37}$$

and

$$R(x|x') = \left[\frac{\omega(x)}{\omega(x')}\right]^{3/2}. \tag{38}$$

In this notation Eq. (32) becomes

$$I_T(x,\mu) = \Theta(x)\Theta(x_b - x)$$
$$= \alpha(x)\Big[R(x|0)A(x,\mu|0) - R(x|x_b)A(x,\mu|x_b) + \int_0^{x_b} dx' R(x|x')B(x,\mu|x')\Big] \tag{39}$$

The corresponding integral equations for the boundary values of $I(x,\mu)$ are obtained by letting $x \to 0, x_b$. These are

$$I_T(0,\mu) = \alpha(0)\Big[A(0,\mu|0) - R(0|x_b)A(0,\mu|x_b) + \int_0^{x_b} dx' R(0|x')B(0,\mu|x')\Big] \tag{40}$$

and

$$I_T(x_b,\mu) = \alpha(x_b)\Big[R(x_b|0)A(x_b,\mu|0) - A(x_b,\mu|x_b) + \int_0^{x_b} dx' R(x_b|x')B(x_b,\mu|x')\Big]. \tag{41}$$

For the strict albedo problem, the boundary conditions are

$$I(0,\mu) = \delta(\mu - \mu_1), \quad \mu,\mu_1 > 0, \tag{42}$$

and

$$I(x_b,\mu) = I_b(x_b,\mu), \quad \mu < 0, \tag{43}$$

where $I_b(x_b,\mu)$ for $\mu < 0$ is the reflected intensity at the bottom. Thus, in principle, one should solve integral equations (40) and (41) under the boundary conditions (42) and (43) to obtain $I_T(0,\mu)$ and $I_T(x_b,\mu)$ for all values of μ. Then knowing A and B from (33)–(35), we determine the state of radiation everywhere from Eq. (39). However, as we mentioned earlier in Sec. 2 that the rigorous solution of the coupled integral equations (40) and (41) is not known, we propose the use of experimental measurements to circumvent that difficulty. We discuss two cases.

A. $I(0,\mu)$ and $I(x_b,\mu)$ given by experiment

With the modern advent of satellites, it is quite feasible to have satellite-borne optical sensors to measure the radiation field of any given frequency at high altitudes to obtain the data on $I(0,\mu)$ for all angles. It then would be required that simultaneous measurement on the ground be made to obtain $I(x_b,\mu)$ in the same meridian plane which contains the sun, the satellite, and the ground-based station. Assuming that we are given both boundary values of $I(x,\mu)$, we compute A and B. Thus, from (33) and (23) we have

$$A(x,\mu|0) = \frac{1}{N_0(x)}\phi_0(\mu,x)e^{-x/\nu_0(x)}a_0(0) + \int_0^1 \frac{d\nu}{N(\nu,x)} e^{-x/\nu}\phi_\nu(\mu,x)a_\nu(0), \tag{44}$$

where

$$a_0(0) = \int_{-1}^1 d\mu\mu I(0,\mu)\phi_0(\mu,0) \tag{45}$$

and

$$a_\nu(0) = \int_{-1}^1 d\mu\mu I(0,\mu)\phi_\nu(\mu,0). \tag{46}$$

The eigenfunctions $\phi_0(\mu,0)$ and $\phi_\nu(\mu,0)$ are defined by (24) and (25), respectively. Similarly from (34) and (23) we have

$$A(x,\mu|x_b) = -\frac{1}{N_{0^-}(x)}\phi_{0^-}(\mu,x)\exp\left(\frac{-(x_b - x)}{\nu_0(x)}\right) - \int_0^1 \frac{d\nu}{N(-\nu,x)}\exp\left(-\frac{(x_b - x)}{\nu}\right) \times \phi_{-\nu}(\mu,x)a_{-\nu}(x_b), \tag{47}$$

where

$$a_{0^-}(x_b) = \int_{-1}^1 d\mu\mu I(x_b,\mu)\,\phi_{0^-}(\mu,x_b) \tag{48}$$

and

$$a_{-\nu}(x_b) = \int_{-1}^1 d\mu\mu I(x_b,\mu)\,\phi_{-\nu}(\mu,x_b). \tag{49}$$

Using (44) and (47) in (39), we obtain

$$I_T(x,\mu)\Theta(x)\Theta(x_b - x) = \alpha(x)\Big\{R(x|0)\Big[\frac{1}{N_0(x)}\phi_0(\mu,x)\exp\Big(-\frac{x}{\nu_0(x)}\Big)a_0(0) + \int_0^1 \frac{d\nu}{N(\nu,x)}e^{-x/\nu}\phi_\nu(\mu,x)a_\nu(0)\Big] + R(x|x_b)\Big[\frac{1}{N_{0^-}(x)}\phi_{0^-}(\mu,x)\exp\Big(-\frac{(x_b - x)}{\nu_0(x)}\Big) + \int_0^1 \frac{d\nu}{N(-\nu,x)}\exp\Big(-\frac{(x_b - x)}{\nu}\Big)\phi_{-\nu}(\mu,x) \times a_{-\nu}(x_b)\Big] + \int_0^{x_b} dx' R(x|x')B(x,\mu|x')\Big\}. \tag{50}$$

Equation (50) provides the state of radiation everywhere in the medium. We have always assumed that the internal radiation source $Q(x,\mu)$ (if any) is always given so that $B(x,\mu|x')$, defined by (35), is known.

B. $I(x_b,\mu)$ given by experiment

If we have the ground based measurement so that $I(x_b,\mu)$ is known for all angles μ, and in addition if we know the incident flux at the top of the atmosphere, then in that case the only unknown quantities in Eq. (50) are $a_0(0)$ and $a_\nu(0)$. These quantities are readily obtained by using the

half-range orthogonality properties of Case's eigenfunctions.[1]

To see that, first rewrite Eq. (50) as follows:

$$I_T(x,\mu)\Theta(x)\Theta(x_b-x)$$
$$=\alpha(x)R(x|0)\Big[\frac{1}{N_0(x)}\phi_0(\mu,x)\exp\Big(-\frac{x}{\nu_0(x)}\Big)a_0(0)$$
$$+\int_0^1\frac{d\nu}{N(\nu,x)}\exp\Big(-\frac{x}{\nu}\Big)\phi_\nu(\mu,x)a_\nu(0)\Big]+F(x,\mu), \quad (51)$$

where

$$F(x,\mu)=\alpha(x)R(x|x_b)\Big[\frac{a_{0-}(x_b)}{N_{0-}(x)}\phi_{0-}(\mu,x)$$
$$\times\exp\Big(-\frac{(x_b-x)}{\nu_0(x)}\Big)+\int_0^1\frac{d\nu}{N(-\nu,x)}$$
$$\times\exp\Big(-\frac{(x_b-x)}{\nu}\Big)\phi_{-\nu}(\mu,x)a_{-\nu}(x_b)\Big]$$
$$+\alpha(x)\int_0^{x_b}dx'R(x|x')B(x,\mu|x'). \quad (52)$$

Now let $x\to 0$ in Eq. (51) and impose the boundary condition (42) for the incident flux. We get

$$\delta(\mu-\mu_1)=\alpha(0)\Big[\frac{1}{N_0(0)}\phi_0(\mu,0)a_0(0)+\int_0^1\frac{d\nu}{N(\nu,0)}$$
$$\phi_\nu(\mu,0)a_\nu(0)\Big]+F(0,\mu), \quad \mu,\mu_1>0. \quad (53)$$

We use the following half-range orthogonality properties of $\phi_0(\mu,0)$ and $\phi_\nu(\mu,0)$ (cf. Eqs. (34a), (34b), and (34d), Sec. 4 in Ref. 1]:

$$\int_0^1 d\mu\,\phi_\nu(\mu,0)\phi_{\nu'}(\mu,0)W(\mu)=W(\nu)\frac{N(\nu,0)}{\nu}\delta(\nu-\nu'), \quad (54)$$

$$\int_0^1 d\mu\,\phi_\nu(\mu,0)\phi_0(\mu,0)W(\mu)=0, \quad (55)$$

$$\int_0^1 d\mu\,\phi_0^2(\mu,0)W(\mu)=-[\tfrac{1}{2}\omega(0)\nu_0(0)]^2X(\nu_0(0)), \quad (56)$$

$$W(\mu)=\gamma(\mu)[\nu_0(0)-\mu], \quad (57)$$

$$\gamma(\mu)=\frac{\omega(0)}{2}\mu\frac{X^+(\mu)}{\Lambda^+(\mu,0)}, \quad (58)$$

$$\Lambda(z,0)=1-\omega(0)\frac{z}{2}\int_{-1}^1\frac{d\mu}{z-\mu}, \quad (59)$$

$$X(z)=\frac{1}{1-z}\cdot\exp\Big[\frac{1}{2\pi i}\int_0^1\frac{d\mu}{\mu-z}\ln\Big(\frac{\Lambda^+(\mu)}{\Lambda^-(\mu)}\Big)\Big] \quad (60)$$

and $+$ on $X^+(\mu)$ and $\Lambda^+(\mu,0)$ are the boundary values of $X(z)$ and $\Lambda(z,0)$ as z approaches their branch cut from the top, respectively. Using the orthogonality properties (54), (55), and (56) in Eq. (53), we obtain

$$a_0(0)=-\frac{N_0(0)}{\alpha(0)X(\nu_0(0))}\Big(\frac{2}{\omega(0)\nu_0(0)}\Big)^2$$
$$\times\Big[\phi_0(\mu_1,0)W(\mu_1)-\int_0^1 d\mu\,F(0,\mu)\phi_0(\mu,0)W(\mu)\Big]$$
and (61)

$$a_\nu(0)=\frac{\nu}{\alpha(0)W(\nu)}\Big[\phi_\nu(\mu_1,0)W(\mu_1)$$
$$-\int_0^1 d\mu\,F(0,\mu)\phi_\nu(\mu,0)W(\mu)\Big]. \quad (62)$$

These coefficients in conjunction with Eq. (51) provide a complete description of the radiation field everywhere.

C. Source free semi-infinite medium

For a source free semi-infinite medium, $F(x,\mu)$ defined by (52) is identically zero, in which case Eq. (51) reduces

$$I_T(x,\mu)\Theta(x)=\alpha(x)R(x|0)$$
$$\times\Big[\frac{1}{N_0(x)}\phi_0(\mu,x)\exp\Big(-\frac{x}{\nu_0(x)}\Big)a_0(0)$$
$$+\int_0^1\frac{d\nu}{N(\nu,x)}\exp\Big(-\frac{x}{\nu}\Big)\phi_\nu(\mu,x)a_\nu(0)\Big]. \quad (63)$$

The coefficients $a_0(0)$ and $a_\nu(0)$, given by (61) and (62) respectively, simplify to

$$a_0(0)=-\frac{N_0(0)}{\alpha(0)X(\nu_0(0))}\Big(\frac{2}{\omega(0)\nu_0(0)}\Big)^2\phi_0(\mu_1,0)W(\mu_1) \quad (64)$$

and

$$a_\nu(0)=\frac{\nu}{\alpha(0)W(\nu)}\phi_\nu(\mu_1,0)W(\mu_1). \quad (65)$$

ACKNOWLEDGMENTS

I am grateful to Professor K.M. Case of Rockefeller University, Professor H.E. Moses of the University of Lowell, and Professor J. Davies of Clark University for reading the manuscript and for making helpful suggestions.

[1] K.M. Case and P.F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[2] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[3] V.A. Ambarzumian, C.R. (Doklady) Acad. Sci. VSSR **38**, 229 (1943).
[4] D.H. Menzel, *Selected Papers on the Transfer of Radiation* (Dover, New York, 1966).
[5] M. Kanal and H.E. Moses, *A variational principle for transport theory*, J. Math. Phys. **19**, 1258 (1978).
[6] M. Kanal, "Boundary-Value Problems of Linear Transport Theory—Green's Function Approach," J. Math. Phys. **11**, 3042 (1970).

Classical oscillators excited parametrically and their applications in plasma physics

S. J. Lucena
Coppe Nuclear, Universidade Federal do Rio de Janeiro, Brasil

T. A. Minelli and M. Pusterla[a)]
Istituto di Fisica, Università di Padova, Italy
and I.N.F.N. Sez, Padova, Italy

(Received 26 February 1979)

Oscillators coupled directly and through time dependent external periodic forces, are considered here as a scheme that describes the motions of two plasmas in interaction with each other and with external fields. The analytical aspects of the equations are investigated both in their differential and integral forms. The formalism of Green's functions is widely emphasized with a particular attention towards the resolvent. "Structural" Green's functions in relation with the periodic time dependence of the external forces are introduced as a valid calculational tool in order to compute relevant physical features of the system. In particular, correlations between the mathematical singularities of the resolvent "structural" Green's function, the dispersion relations, and the parametric instabilities are established. Exact solutions, with the use of the Green's function approach and the direct one on the amplitudes, are obtained for particular models that exhibit a quite simplified but physically meaningful time dependence for the periodic external forces.

I. INTRODUCTION

Physical systems with a high degree of complexity can often be analyzed, approximately, in terms of coupled oscillators. Such an approach appears particularly promising in plasma physics, where the basic nonlinear equations require a procedure of linearization in order to allow the study of the interactions of the various components with each other and with external systems. The process of linearizing the plasma electromagnetic equations, in fact, as has been shown by many authors,[1-3] leads to a system of oscillators coupled through a field which may be external (pump field) or internal. Consequently it also leads to the determination of the normal modes, the dispersion relations and to the analysis of the parametric instabilities which constitutes an important goal in the theory of plasmas. These physical points can be achieved either by the use of sophisticated numerical codes or by the introduction of analytical calculational methods that usually provide a much deeper understanding of the physical phenomena but need drastic simplifications in order to make the equations solvable. To be more explicit, if one follows this second line, starting from the basic linearized equations one immediately faces two alternatives: either an analytical approximation of the solutions, obtained for instance by iterative methods, or the replacement of the interaction terms, present in the equations, with new ones that appear simple enough to permit an exact evaluation of the equations, possibly in terms of elementary functions. The first possibility is more systematic and is used in principle, in all cases where an iterative process can be established. One can develop this method either on the physical amplitudes (in the oscillator description) or on the Green's functions in both coordinate and Fourier spaces. The iterative procedure usually leads to a formal expansion of the solution that requires a technique of recovery of the series (we may suggest the Padé approximants[4] or similar algorithms) in order to interpret the approximation physically. Such an approach deserves great attention and will be emphasized in a subsequent paper. In this paper we develop the second point of view mentioned above, namely, the replacement of exact interaction terms in the equations of the coupled oscillators with simplified terms that retain several physical aspects of the previous ones and allow an exact compact solution because of their simplicity. We have in mind the system of two or more oscillators excited by a periodic external field (pump field). In this context the exactly solvable models, which we are introducing, are equivalent, as will be apparent later, to the replacement of a certain sinusoidal function (usually related to the external field) with another one, that is still periodic with the same period and consequently has in addition to the fundamental harmonic the infinite Fourier superposition of the higher harmonics. Their weights must be negligible, if the approximation is to make sense. Such a substitution appears perfectly legitimate and justified, being on the same footing as the standard approach[1,2] that reduces the problem of determining the infinite set of the Fourier components of the physical solution (see Appendix D) to an algebraic system, by cutting the higher harmonics, obviously considering their contributions negligible. The method we are proposing adds higher harmonics instead of neglecting them but has the advantage of giving an exact compact representation of the physical solution. In this paper we are mainly interested in the coupled oscillator system in classical physics and particular emphasis is given to the Green's func-

[a)]Research partially sponsored by C.O.P.P.E. (UFRJ) and C.N. Pq. (Brasil).

 0022-2488/79/122639-10$01.00

tion formalism (Secs. II, III, and IV) whereas Sec. IV covers the direct calculation of the coupled amplitudes.

II. GREEN'S FUNCTIONS FOR COUPLED OSCILLATORS

A. Integral equations for parametric excitations

Let us consider the simplest system having characteristics of interest for our purposes: two oscillators $X_1(t), X_2(t)$ coupled with each other by both conservative and time dependent, nonconservative forces, which are usually called parametric excitations. The equations of motion can be written in a compact way,

$$LX(t) = \mu Z(t) X(t), \tag{1a}$$

where

$$L = \begin{pmatrix} \frac{d^2}{dt^2} + \gamma_1 \frac{d}{dt} + \omega_1^2 & -\lambda_{12} \\ -\lambda_{21} & \frac{d^2}{dt^2} + \gamma_2 \frac{d}{dt} + \omega_2^2 \end{pmatrix},$$

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \tag{1b}$$

(μ is a constant matrix while Z is a numerical function of t).

It is obvious that Eq. (1) represents the extension to the two oscillators (and possibly to a system of three or more oscillators) of the equation of motion of one oscillator in the interaction with itself:

$$\ddot{X} + \gamma \dot{X} + \tilde{\omega}_0^2 X = \lambda X + \mu Z(t) X. \tag{2}$$

For such an equation we know how the Green's function $G_0(t,t')$ containing the desired boundary conditions must be built. More specifically, if we like solutions that exhibit the asymptotic conditions for the t variable analogous to the Sommerfeld radiation ones,[5] $G_0(t,t')$ acquires a very simple expression in terms of the two linearly independent solutions X_1, X_2, of the homogeneous equation

$$\ddot{X} + \gamma \dot{X} + \omega_0^2 X = 0, \quad \omega_0^2 = \tilde{\omega}_0^2 - \lambda, \tag{3}$$

$$X_1(t) = e^{\alpha t}, \quad X_2(t) = e^{\alpha^* t}, \tag{4}$$

where α, α^* are the two complex conjugate roots of the second degree (real coefficients) equations $Z^2 + \gamma Z + \omega_0^2 = 0$.

We have

$$G_0(t,t') = \{1/W[X_1(t');X_2(t')]\} X_1(t) X_2(t')$$
$$= \frac{e^{\alpha(t-t')}}{2i\Omega} \quad (\text{for } t < t'), \tag{5a}$$

$$G_0(t,t') = \{1/W[X_1(t');X_2(t')]\} X_1(t') X_2(t)$$
$$= \frac{e^{\alpha^*(t-t')}}{2i\Omega} \quad (\text{for } t > t'), \tag{5b}$$

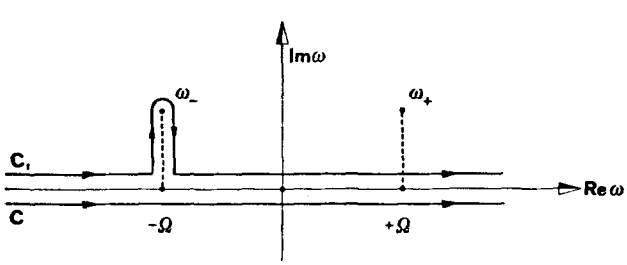

FIG. 1. Contours C_1 and C.

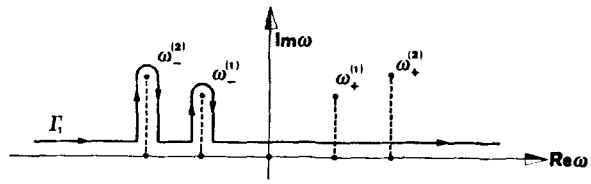

FIG. 2. Contour Γ_1.

$$\alpha = -\tfrac{1}{2}\gamma - i\Omega, \quad \Omega = (\omega_0^2 - \gamma^2/4)^{1/2}. \tag{5c}$$

$W[X_1(t');X_2(t')]$ is the Wronksian built with the two solutions $X_1(t'), X_2(t')$.

Equation (2), together with the asymptotic conditions, is therefore equivalent to the following integral equation:

$$X(t) = \mu \int_{-\infty}^{\infty} G_0(t,t') Z(t') X(t')\, dt'. \tag{6}$$

$G_0(t,t')$ is simply obtained by choosing the contour C_1 in the Fourier integral of the function $(-\omega^2 + i\gamma\omega + \omega_0^2)^{-1}$, the poles of which are $\omega_\pm = i(\gamma/2) \pm \Omega$:

$$G_0(t,t') = (2\pi)^{-1} \int_{C_1} d\omega\, e^{i\omega(t-t')} \times [-\omega^2 + i\gamma\omega + \omega_0^2]^{-1}. \tag{7}$$

We should like to point out that the integral representation used in Eq. (7) differs from the usual prescription that introduces another contour $C \neq C_1$, defined along the real axis, and generates the so-called one point boundary condition Green's function.[6]

We now come to Eq. (1), written in a compact matrix form, for two (or more than two) oscillators. It is equivalent to the integral equation

$$X(t) = \int_{-\infty}^{\infty} G_0(t,t') \mu Z(t') X(t')\, dt', \tag{8}$$

where the 2×2 matrix $G_0(t,t')$, with the property of providing the required boundary conditions, must be obtained by a Fourier integration of the matrix, function of ω,

$$\begin{pmatrix} -\omega^2 + i\gamma_1\omega + \omega_1^2 & -\lambda_{12} \\ -\lambda_{21} & -\omega^2 + i\gamma_2\omega + \omega_2^2 \end{pmatrix}^{-1}$$

along a contour that tends to C_1 in the limit of decoupled oscillators ($\lambda_{12}, \lambda_{21} \to 0$). This contour is Γ_1 of Fig. 2.

The poles $\omega_\pm^{(1)}, \omega_\pm^{(2)}$ are the complex roots of the fourth degree equation

$$\omega^4 - i(\gamma_1 + \gamma_2)\omega^3 - (\omega_1^2 + \omega_2^2 + \gamma_1\gamma_2)\omega^2 + i(\gamma_2\omega_1^2 + \gamma_1\omega_2^2)\omega + \omega_1^2\omega_2^2 - \lambda_{12}\lambda_{21} = 0. \tag{9}$$

The roots of Eq. (9) and their locations in the complex ω-plane can easily be analyzed passing to the variable $\tilde{\omega} = -i\omega$. It has to verify a fourth degree equation with real coefficients; in fact, we can take their algebraic representation:

$$\tilde{\omega}_+^{(i)} = \tfrac{1}{4}\left[-(b \pm 2P) \mp \sqrt{(b \pm 2P)^2 - 8(s \pm 2Q)}\,\right]$$
(i labels the inner bracket sign) (10)

where the symbols b, P, Q, s are defined in Appendix A.

One should notice that s is a root of the third degree resolvent equation with real coefficients and can always be taken as real.

As is shown in detail in Appendix A, we can draw the conclusion that the four roots $\omega_{\pm}^{(i)}$ $(i=1,2)$, lie on the same half (part) of the ω complex plane [with our conventions $\mathrm{Im}(\omega_{\pm}^{(i)})>0$], symmetrically located with respect to the imaginary axis, as a consequence of the following obvious conditions that must be verified:

(a) positivity of the damping parameters γ_1, γ_2;

(b) possibility of obtaining the given normal modes of the two independent oscillators in the decoupling limit $(\lambda_{12},\lambda_{21}\to 0)$;

(c) possibility of considering the zero damping limit, that makes sense for strong coupling and high frequencies.

The 2×2 matricial Green's function $G_0(t,t')$ can now be written simply:

$$G_0(t,t') = \Theta(t-t')\left[e^{i\omega_+^{(1)}(t-t')}M_1^+ + e^{i\omega_+^{(2)}(t-t')}M_2^+\right] + \Theta(t'-t)\left[e^{i\omega_-^{(1)}(t-t')}M_1^- + e^{i\omega_-^{(2)}(t-t')}M_2^+\right], \tag{11a}$$

where

$$M_h^{\pm} = \frac{\pm i}{(\omega_{\pm}^{(h)}-\omega_{\mp}^{(h)})(\omega_{\pm}^{(h)}-\omega_{-}^{(k)})(\omega_{\pm}^{(h)}-\omega_{+}^{(k)})}\begin{pmatrix} -(\omega_{\pm}^{(h)})^2+2\gamma_2\omega_{\pm}^{(h)}+\omega_2^2 & \lambda_{12} \\ \lambda_{21} & -(\omega_{\pm}^{(h)})^2+i\gamma_1\omega_{\pm}^{(h)}+\omega_1^2 \end{pmatrix}, \tag{11b}$$

with $h=1,2$; $k=1,2$; $h\neq k$.

It is easy to perform the extension to a system of more than two oscillators, because one expects the properties for the contour of integration and the matrix elements of G_0 to hold in general.

B. Resolvent Green's function

The relation between $G_0(t,t')$ and the resolvent $G(t,t')$ follows from the formal identity

$$G = G_0 + G_0 0 G \quad \text{with } 0 = G_0^{-1} - G^{-1}, \tag{12a}$$

namely,

$$G(t,t') = G_0(t,t') + \int_{-\infty}^{\infty} G_0(t,t'')Z(t'') \times \mu G(t'',t')\,dt''. \tag{12b}$$

III. PERIODIC EXCITATIONS AND STRUCTURAL GREEN'S FUNCTIONS

A. Periodic excitations

We are particularly interested in the case where $Z(t)$ is a periodic function of time with a certain period T, having in mind the physical model of Nishikawa[1] and instabilities at the critical surface of the inertially confined plasma as it develops from an irradiation with a powerful laser.[2] The system of Eqs. (1) becomes a special case of the theory of linear differential equations with periodic coefficients that have the so-called Floquet solutions characterized by the behavior[7]

$$X(t+T) = \lambda X(t). \tag{13}$$

Equation (13) for each fixed value of t_0 of t defines the following boundary conditions[8] in an interval whose extension coincides with a period $T=2\pi/\omega_L$:

$$X(t_0+T) = \lambda X(t_0), \quad \dot{X}(t_0+T) = \lambda \dot{X}(t_0). \tag{14}$$

In correspondence to each value λ_k of the unknown parameter λ allowing such solutions we have a different oscillation mode (and a different kind of wave). Values ω_k of parameter ω, defined by $e^{i\omega T}\equiv\lambda$, give the so-called normal mode frequencies of the system. To be more precise, the real part of the ω_k gives the frequencies (cyclic frequency equal to $2\pi\nu_k$) of the mode, whereas the imaginary part of ω_k gives the damping or the growing rate, according to its sign.[9]

It is obvious that each fundamental normal mode frequency is accompanied by the infinite sequence $\omega_{kn}=\omega_k+n\omega_L$. From Eq. (13) we can define $X(t)$ in the interval $[-T/2,T/2]$ and extend it to the whole real axis. As an immediate consequence we have the possibility of substituting the integral equations (6), (8) with others exhibiting a different Green's function, the so-called structural Green's function (SGF) $G_0^S(\omega,t,t')$ with boundary conditions of the type introduced above. This can be obtained easily by breaking the integration domain into parts each having the same extension, one period T, and shifting all the integration intervals to the only one going from $-T/2$ to $T/2$, with the help of Eq. (13) and the periodicity of $Z(t)$.[10]

So we have

$$X(t) = \int_{-T/2}^{T/2} G_0^S(\omega,t,t')\mu Z(t')X(t')\,dt', \tag{15a}$$

where

$$G_0^S(\omega,t,t') = \sum_{n=-\infty}^{+\infty} G_0(t;t'+nT)e^{in\omega T}. \tag{15b}$$

After simple explicit calculations we obtain for the one dimensional case [Eq. (6)]:

$$\begin{aligned} &G_0^S(\omega,t,t') \\ &= \Theta(t-t')\{g(i\omega-\alpha)e^{\alpha(t-t')} + [1+g(\alpha^*-i\omega)] \\ &\quad\times e^{\alpha^*(t-t')}\} + \Theta(t'-t)\{[1+g(i\omega-\alpha)]e^{\alpha(t-t')} \\ &\quad + g(\alpha^*-i\omega)e^{\alpha^*(t-t')}\}, \end{aligned}$$

$$\text{where } g(z) = \sum_{1}^{\infty} e^{nTz} = (e^{-zT}-1)^{-1}, \tag{16}$$

and for the 2×2 matricial case [Eq. (18)]:

$$\begin{aligned} G_0^S(\omega,t,t') = G_0(t,t') &+ g[i(\omega-\omega_+^{(1)})]e^{i\omega_+^{(1)}(t-t')}M_1^+ \\ &+ g[i(\omega-\omega_+^{(2)})]e^{i\omega_+^{(2)}(t-t')}M_2^+ \\ &+ g[-i(\omega-\omega_-^{(1)})]e^{i\omega_-^{(1)}(t-t')}M_1^- \\ &+ g[-i(\omega-\omega_-^{(2)})]e^{i\omega_-^{(2)}(t-t')}M_2^-. \end{aligned} \tag{17}$$

B. Resolvent SGF

In correspondence to the structural Green's function G_0^S one defines the resolvent[11] G^S which verifies the nonhomogeneous equation

$$G^S = G_0^S + G_0^S \tilde{0} G^S, \tag{18}$$

when $\tilde{0}$ is the interaction term defined in the unit periodic interval; in our case

$$G^S(t,t') = G_0^S(t,t') + \int_{-T/2}^{T/2} G_0^S(t,t'')Z(t'') \times \mu G^S(t'',t')\,dt''. \tag{19}$$

From the standard theory of integral equations we easily derive the statement that the values of ω, for which the homogeneous Eq. (15a) has nontrivial solutions, namely, the frequencies of the basic modes and their side band harmonics, appear as poles of the resolvent $G^S(\omega,t,t')$.

Its residue ρ_n can be expressed in terms of eigenfunctions $X_n(t)$, $\tilde{X}_n(t)$ of (15) and of the adjoint equation. More precisely, for simple poles one has

$$\rho_n = \text{const} X_n(t)\tilde{X}_n(t'). \tag{20}$$

C. SGF in a laser fusion model

A physical case, which can be treated in a similar manner, is the interaction between the electron plasma near the critical density surface (as is clearly shown by Brueckner and Jorna[2]) in the evolution of the D–T pellet in laser fusion. In fact, the laser field can be treated as spatially homogeneous (dipole approximation) and one deduces, from the basic general equations the following system (see Appendix B for symbols undefined here):

$$\ddot{\tilde{n}}_e + 2\Gamma_e\dot{\tilde{n}}_e + \tilde{n}_e(\omega_e^2 + k^2S_e^2) = \omega_e^2\tilde{n}_i + [-2i\mathbf{k}\cdot\mathbf{V}_0(t)]\dot{\tilde{n}}_e + \tilde{n}_e[-i\mathbf{k}\cdot\dot{\mathbf{V}}_0(t) - 2i\mathbf{k}\cdot\mathbf{V}_0(t)\Gamma_e - (i\mathbf{k}\cdot\mathbf{V}_0)^2], \tag{21a}$$

$$\ddot{\tilde{n}}_i + 2\Gamma_i\dot{\tilde{n}}_i + (\omega_i^2 + k^2S_i^2)\tilde{n}_i = \omega_i^2\tilde{n}_e, \tag{21b}$$

where

$$\tilde{n}_\mu(\mathbf{k},t) = (2\pi)^{-3/2}\int d^3\mathbf{r} n_\mu(\mathbf{r},t)e^{-i\mathbf{k}\cdot\mathbf{r}},$$

$n_\mu(\mathbf{r},t)$ being the density of the μ plasma; $\mathbf{V}_0(t)$ is related with the linearly polarized laser field $\mathbf{E}$ and $\mathbf{H}$ and in our approximations ($\mathbf{k}_L \simeq 0$), $\mathbf{V}_0(t) = \tilde{\mathbf{V}}_0 \sin\omega_L t$; $\mathbf{E}(t) = \mathbf{E}_0\cos\omega_L t$; $\tilde{\mathbf{V}}_0 = -e\mathbf{E}_0/m_e\omega_L$.

Equations (21), although more complicated, can be analyzed by our 2×2 matrix formalism, the only difference being that the interaction operator has a more complex structure:

$$Z(t)\mu = \begin{pmatrix} \alpha(t)\dfrac{d}{dt} + \beta(t) & 0 \\ 0 & 0 \end{pmatrix},$$

where

$$\alpha(t) = -2i\mathbf{k}\cdot\mathbf{V}_0(t),$$

$$\beta(t) = -i\mathbf{k}\cdot\mathbf{V}_0 - 2i\mathbf{k}\cdot\mathbf{V}_0\Gamma_e - (i\mathbf{k}\cdot\mathbf{V}_0)^2,$$

or

$$\alpha(t) = -2i\mathbf{k}\cdot\mathbf{V}_0\sin\omega_L t,$$

$$\beta(t) = -i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\omega_L\cos\omega_L t - 2i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\Gamma_e\sin\omega_L t + (\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2\sin^2\omega_L t.$$

The Green's function $G^S(t,t')$ verifies the integral equation:

$$G^S(t,t') = G_0^S(t,t') + \int_{-T/2}^{T/2} dt''\, G_0^S(t,t'') \times\left(\alpha(t'')\frac{d}{dt''} + \beta(t'')\right)\zeta_+ G^S(t'',t'), \tag{22}$$

where

$$\zeta_+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

If one integrates by parts, Eq. (22) can be reduced to the following one:

$$G^S(t,t') = G_0^S(t,t') + \int_{-T/2}^{T/2} dt'' \times\left[\left(-\frac{d}{dt''}\alpha(t'') + \beta(t'')\right)G_0^S(t,t'')\right] \times\zeta_+ G^S(t'',t'). \tag{23}$$

IV. KRONIG–PENNEY MODELS OF PARAMETRIC EXCITATION

A. Exactly solvable models

Equations (19), (23), being inhomogeneous, can be solved by iterative methods; the expansion, however, could be divergent or slowly converging and the search for the singularities of the resolvent SGF usually requires the summation of the series, at least approximately, as one can do by various algorithms such as Padé approximants[4] or similar methods. The alternative possibility lies in evaluating those equations exactly, obtaining the resolvent G^S in an exact analytically closed form, after one has reformulated the functional dependence on t of the interaction terms. In this case, by redefining the interaction terms, one makes an approximation at the initial stage of the calculation of G^S.

We are going to emphasize such an alternative here and in the coming section where we deal with the amplitudes, leaving the iterative methods to a separate communication. The resolvent G^S can be given a compact representation as an exact solution of Eq. (19) [similarly, we can deal with Eq. (22) as will be shown explicitly] if the realistic periodic function $Z(t)$ is replaced by another one, $\tilde{Z}(t)$, that keeps some peculiar features of the more physically correct $Z(t)$: one chooses a periodical, properly weighted sequence of square wells or their limiting case, delta functions.

The delta function sequence is treated here, the square wells in the next section. Further details are developed in Apppendix C. It is interesting to note at this point that our approach in classical coupled oscillators has an analogue correspondence in solid state physics where periodic potentials can be successfully approximated by a periodic sequence of square wells or delta functions[12] (Kronig–Penney models).

B. Exact representation of G^S with particular functions $Z(t)$

To be more explicit, instead of the realistic excitation $Z(t) = Z_0\sin\omega_L t$, entering the Nishikawa model, we take the Kronig–Penney extreme model (see Appendix C):

$$\tilde{Z}(t) = \sum_{-\infty}^{\infty}{}_n \; [\delta(t-(2n+1)T/4)](-1)^n \Gamma Z_0, \qquad \Gamma = 2/\omega_L.$$

This choice leads, from Eq. (19), to the following representation:

$$G^S(t,t') = G_0^S(t,t') + \Gamma\,[G_0^S(t,T/4)\mu G^S(T/4,t') - G_0^S(t,-T/4)\mu G^S(-T/4,t')], \tag{24}$$

where $G^S(T/4,t')$ and $G^S(-T/4,t')$ can be obtained from the equation

$$\begin{pmatrix} -\Gamma G_0^S(T/4,T/4)\mu + 1 & \Gamma G_0^S(T/4,-T/4)\mu \\ -\Gamma G_0^S(-T/4,T/4)\mu & \Gamma G_0^S(-T/4,-T/4)\mu + 1 \end{pmatrix} \times \begin{pmatrix} G^S(T/4,t') \\ G^S(-T/4,t') \end{pmatrix} = \begin{pmatrix} G_0^S(T/4,t') \\ G_0^S(-T/4,t') \end{pmatrix}. \tag{25}$$

It follows immediately, from (24) and (25), that the "normal mode frequencies" of the system are given by the determinantal equation

$$\det \begin{pmatrix} 1-\Gamma G_0^S(T/4,T/4)\mu & \Gamma G_0^S(T/4,-T/4)\mu \\ -\Gamma G_0^S(-T/4,T/4)\mu & \Gamma G_0^S(-T/4,-T/4)\mu \end{pmatrix} = 0. \tag{26}$$

For the system of two oscillators the determinant of Eq. (26) is 4×4.

In a more cumbersome but completely analogous way one starts from Eqs. (22) and (23) and considers the periodic functions $\tilde{\alpha}(t)$, $\tilde{\beta}(t)$ defined by periodic sequences of deltas instead of the sinusoidal ones (see Appendix C for details). One arrives easily at the following representation for the resolvent $G^S(t,t')$:

$$G_{ij}^S(t,t') = G_{0\,ij}^S(t,t') + \sum_{1}^{5}{}_k \, S_{i1}^k(t,t_k) G_{1j}^S(t_k,t'), \tag{27}$$

where $S^k(t,t_k)$ and $G^S(t_k,t')$ can be found in Appendix C.

This allows an easy derivation of the dispersion relations in a determinantal form:

$$\det A = 0, \tag{28}$$

where

$$A_{ij} = [S_{11}^i(t_j,t_i) - \delta_{ij}].$$

C. Direct method

In this subsection we focus our attention on the problem of solving directly the coupled oscillator equations in their differential form when coupling is periodic.

This procedure can be looked at as alternative to the Green's function method and often permits one to reach the desired solutions more easily.

Let us take, for simplicity, the case of the two oscillators $X_1(t)$, $X_2(t)$; their coupled equations can be rewritten as a 4×4 linear system of the first order:

$$\frac{d}{dt}\mathbf{U} = [A_0 + A_1(t)]\mathbf{U}, \tag{29a}$$

with

$$U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} = \begin{pmatrix} X_1 \\ \dot{X}_1 \\ X_2 \\ \dot{X}_2 \end{pmatrix},$$

$$A_0 \equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\omega_1^2 & -\gamma_1 & \lambda_{12} & 0 \\ 0 & 0 & 0 & 1 \\ \lambda_{21} & 0 & -\omega_2^2 & -\gamma_2 \end{pmatrix}, \tag{29b}$$

$$A_1(t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{11}(t) & \alpha(t) & \beta_{12}(t) & 0 \\ 0 & 0 & 0 & 0 \\ \beta_{21}(t) & 0 & \beta_{22}(t) & 0 \end{pmatrix},$$

where $\alpha = 0, \beta_{ij} = Z(t)\mu$ for the Nishikawa model while $\beta_{11} = \beta, \beta_{12} = \beta_{21} = \beta_{22} = 0$ for the linearized model of laser fusion.

The differential system (29) has a general solution of the type $\mathbf{U}(t) = \Sigma_{1\,m}^4 C_m \mathbf{U}^{(m)}$ where the $\mathbf{U}^{(m)}$'s are four particular linearly independent solutions of the same system of equations. If one introduces the square matrix $U_{nm}(t) = U_n^{(m)}(t)$, called the fundamental matrix, where n labels the nth component of the column vector $\mathbf{U}^{(m)}$, the linear independence is expressed by the condition that the determinant (Wronksian) of the square matrix U_{nm} is different from zero.

We now impose the boundary conditions (14), in agreement with the arguments of Sec. IIIA, on the general solution $\mathbf{U}(t)$

$$\left[\sum_{1}^{4}{}_m \mathbf{U}^{(m)} C_m\right]_{t=t_0+T} = \lambda \left[\sum_{1}^{4}{}_m \mathbf{U}^{(m)} C_m\right]_{t=t_0} \tag{30}$$

or

$$\sum_{1}^{4}{}_m \, [U_{nm}(t_0+T) - \lambda U_{nm}(t_0)] C_m = 0 \tag{31}$$

(with $n = 1,2,3,4$).

This set of four homogeneous equations requires the determinantal condition

$$\det[U(t_0+T) - \lambda U(t_0)] = 0, \tag{32}$$

which is the eigenvalue equations for λ.

The fundamental matrix U is called principal if $U_{nm}(t_0) = \delta_{nm}$; we can always refer to this simplified case reducing the secular equations to the form

$$\det[U(t_0+T) - \lambda I] = 0. \tag{33}$$

Equation (32), or its simplified version (33), is naturally hard to solve because of the difficulty of obtaining four linearly independent integrals $\mathbf{U}^{(m)}(t)$'s once the interaction matrix $A_1(t)$ is assigned. However, it permits us, in principle, to determine the four eigenvalues λ_k ($k = 1,2,3,4$) and the corresponding frequencies ω_k (in general complex) of the normal modes. It is therefore at this point that we may replace the exact matrix $A_1(t)$, or more specifically, the functions $Z(t), \alpha(t), \beta(t)$, with periodic sequences of square wells $\tilde{Z}(t), \tilde{\alpha}(t), \tilde{\beta}(t)$ that have their depths and widths correlated in a definite way with the originals, as is illustrated by Appendix C, reducing our study to one periodicity interval (we

take $[-T/2, +T/2]$) in agreement with what we have pointed out in the preceding subsection. The square-well functions (similar arguments hold for the delta functions) allow us an exact determination of the desired physical amplitudes U's. In fact, we divide the interval $[-T/2, +T/2]$ into parts, where the *nth* subinterval can be written $[nd,(n+1)d]$ with $n = -2,0$ and $d = T/4$ in our first physical case and $n = -4,\ldots,+3$ and $d = T/8$ in the second; within the nth subinterval $A_1(t)$ is a constant matrix $A_1^{(n)}$ and the differential system (29a) takes the form

$$\frac{d}{dt}\mathbf{U}_{(n)} = [A_0 + A_1^{(n)}]\mathbf{U}_{(n)}. \tag{34}$$

It is then simple to search for solutions $\mathbf{U}_{(n)} = e^{\lambda^{(n)}t}\mathbf{V}_{(n)}$ and one deals with the matrix equation

$$[A_0 + A_1^{(n)}]\mathbf{V}_{(n)} = \lambda^{(n)}\mathbf{V}_{(n)}. \tag{35}$$

We then have four eigenvalues $\lambda_k^{(n)}$ $(k = 1,\ldots,4)$ and four eigenvectors $\mathbf{V}_{(n)}$; correspondingly we get four linearly independent functions $\mathbf{U}_{(n)}^k = e^{\lambda k^{(n)}t}\mathbf{V}_{(n)}^k$. The general solution in the nth subinterval is then

$$\mathbf{U}_{(n)} = \sum_1^4 {}_k C_k^{(n)}\mathbf{U}_{(n)}^k. \tag{36}$$

Let us consider the first subinterval, which obviously has $-T/2$ as a lower extreme value and corresponds to a given integer $\bar{n}$; then we define the principal matrix $U_{ij} = U_{(\bar{n})i}^j$ by the usual condition $U_{ij}(t_0 \equiv -T/2) = \delta_{ij}$. This fixes the four constants $C_k^{(\bar{n})}$. The continuity condition in the t-variable for the matrix elements U_{ij} is achieved by imposing the following relations at the border of the $(n-1)$th and the nth interval:

$$\begin{aligned} U_{ij}(nl^+) &= U_{(n)i}^j(nl^+) = U_{(n-1)i}^j(nl^-) \\ &= U_{ij}(nl^-). \end{aligned} \tag{37}$$

Equations (37) extend the principal matrix to the whole interval $[-T/2, +T/2]$ by the definition of the $C_k^{(m)}$'s in the nth subinterval. Once the principal matrix is obtained in the whole interval $[-T/2,T/2]$ the calculation of the normal mode frequencies of the amplitudes follows from the general procedure summarized by Eqs. (31) and (33).

ACKNOWLEDGMENTS

One of the authors (M.P.) is grateful to the Coordinator of the COPPE NUCLEAR (UFRJ, Rio de Janeiro, Brasil), Professor Zieli Thomé, for his kind hospitality, to the COPPE General Direction, and CNPq (Brasil) for their partial support.

APPENDIX A

A. Determination of the Green's function $G_0(t,t')$ for the coupled oscillators with damping

Taking the Fourier transform $\hat{f}(\omega) = (2\pi)^{-1/2}$ $\int_{-\infty}^{\infty} dt\, f(t)e^{-i\omega t}$ of system (1) one obtains $\hat{L}\hat{X} = (\Lambda X)\hat{}$ or

$$\begin{pmatrix} -\omega^2 + i\gamma_1\omega + \omega_1^2 & -\lambda_{12} \\ -\lambda_{21} & -\omega^2 + i\gamma_2\omega + \omega_2^2 \end{pmatrix}\hat{X} = (\Lambda X)\hat{}. \tag{A1}$$

Inverting the matrix $\hat{L}$ one has

$$\hat{X} = \frac{1}{\det\hat{L}(\omega)} \times \begin{pmatrix} -\omega^2 + i\gamma_2\omega + \omega_2^2 & \lambda_{12} \\ \lambda_{21} & -\omega^2 + i\gamma_1\omega + \omega_1^2 \end{pmatrix}(\Lambda X)\hat{}, \tag{A2}$$

and then

$$\begin{aligned} X = (2\pi)^{-1/2}&\int_{-\infty}^{\infty} d\omega \frac{e^{i\omega t}}{\det\hat{L}(\omega)} \\ &\times\begin{pmatrix} -\omega^2 + i\gamma_2\omega + \omega_2^2 & \lambda_{12} \\ \lambda_{21} & -\omega^2 + i\gamma_1\omega + \omega_1^2 \end{pmatrix} \\ &\times(2\pi)^{-1/2}\int_{-\infty}^{\infty} dt'\,(\Lambda X)e^{-i\omega t'}. \end{aligned}$$

Therefore, the matricial Green's function

$$\begin{aligned} G_0(t,t') = \frac{1}{2\pi}&\int_{-\infty}^{\infty} d\omega \frac{e^{i\omega(t-t')}}{\det\hat{L}(\omega)} \\ &\times\begin{pmatrix} -\omega^2 + i\gamma_2\omega + \omega_2^2 & \lambda_{12} \\ \lambda_{21} & -\omega^2 + i\gamma_1\omega + \omega_1^2 \end{pmatrix} \end{aligned} \tag{A3}$$

allows us to transform the differential system (A1) into a system of coupled integral equations.

To find the poles of G_0 [roots of $\det\hat{L}(\omega)$], we make the change of variable $\tilde{\omega} = -i\omega$, and our determinant becomes a polynomial with real coefficients in the new variable:

$$\begin{aligned} \det\hat{L}(\omega) &= \det\hat{L}(i\tilde{\omega}) \\ &= \tilde{\omega}^4 - (\gamma_1+\gamma_2)\tilde{\omega}^3 + (\gamma_1\gamma_2 + \omega_1^2 + \omega_2^2)\tilde{\omega}^2 \\ &\quad - (\gamma_2\omega_1^2 + \gamma_1\omega_2^2)\tilde{\omega} + \omega_1^2\omega_2^2 - \lambda_{12}\lambda_{21}, \end{aligned} \tag{A4}$$

with the following definition: $b = -(\gamma_1+\gamma_2)$; $d = -(\gamma_2\omega_1^2 + \gamma_1\omega_2^2)$; $c = (\gamma_1\gamma_2 + \omega_1^2 + \omega_2^2)$; $e = (\omega_1^2\omega_2^2 - \lambda_{12}\lambda_{21})$, the complex conjugate roots of the complete fourth degree polynomial $\tilde{\omega}^4 + b\tilde{\omega}^3 + c\tilde{\omega}^2 + d\tilde{\omega} + e$ are given by the formulas:

$$\tilde{\omega}_{\pm}^{(i)} = \tfrac{1}{4}\left[-(b \pm 2P) \mp \sqrt{(b \pm 2P)^2 - 8(s \pm 2Q)}\,\right], \qquad i = \begin{cases}1\\2,\end{cases} \tag{A5}$$

where

$$P = \left(\frac{b^2}{4} - C + s\right)^{1/2}, \quad Q = \left(\frac{s^2}{4} - e\right)^{1/2},$$

$$s = \left(\frac{-D + \sqrt{R}}{2}\right)^{1/3} + \left(\frac{-D - \sqrt{R}}{2}\right)^{1/3} + \frac{c}{3},$$

$$R = D^2 + \tfrac{4}{27}C^3, \quad C = \tfrac{1}{3}(3bd - 12e - c^2),$$

$$D = \tfrac{1}{27}(-27d^2 - 27b^2e + 72ce + 9bcd - 2c^3),$$

where s is a root of the resolvent third degree equation with real coefficients. Therefore, s can always be taken as real.

(i) No damping limit case: $(\gamma_1 = \gamma_2 = 0)$.

In this case the poles (normal mode frequencies of vibration) are real.

$$\omega_{\pm}^{(i)} = \pm\left[\tfrac{1}{2}(\omega_1^2 + \omega_2^2 \mp \sqrt{(\omega_1^2 - \omega_2^2)^2 + 4\lambda_{12}\lambda_{21}}\,)\right]^{1/2}, \qquad i = \begin{cases}1\\2.\end{cases} \tag{A6}$$

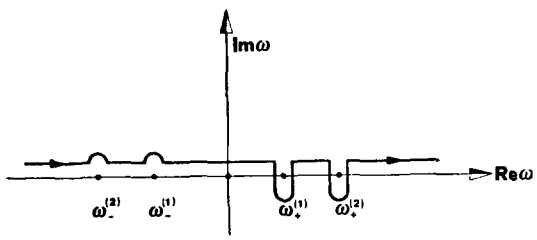

FIG. 3. Contour for the no damping case.

The way of going around the poles was chosen in order to fit the Sommerfeld boundary conditions.

Then

$$G_0(t,t') = \sum_1^2 {}_i(-1)^i \frac{e^{i\omega^{(i)}|t-t'|}}{2i\omega^{(1)}[(\omega^{(1)})^2 - (\omega^{(2)})^2]} \times \begin{pmatrix} -(\omega^{(i)})^2 + \omega_2^2 & \lambda_{12} \\ \lambda_{21} & -(\omega^{(i)})^2 + \omega_1^2 \end{pmatrix}, \tag{A7}$$

where $\omega^{(i)} = |\omega_\pm^{(i)}|$

(ii) Decoupling limit case: $(\lambda_{12} = \lambda_{21} = 0)$

In this case the operator is diagonal, so the oscillators are independent.

For each oscillator, if $\gamma_1 \neq 0$ and $\gamma_2 \neq 0$, the poles are always on the upper half-plane.

Using the same contour of Fig 2 we find

$$G_0(t,t') = \begin{pmatrix} \dfrac{e^{i\Omega_1|t-t'|}e^{-(\gamma_1/2)(t-t')}}{2i\Omega_1} & 0 \\ 0 & \dfrac{e^{i\Omega_2|t-t'|}e^{-(\gamma_2/2)(t-t')}}{2i\Omega_2} \end{pmatrix}, \tag{A8}$$

where

$$\Omega_1 = [\omega_1^2 - (\gamma_1/2)^2]^{1/2}, \quad \Omega_2 = [\omega_2^2 - (\gamma_2/2)^2]^{1/2}.$$

Therefore, in the general case ($\lambda_{12}, \lambda_{21}, \gamma_1, \gamma_2$ all $\neq 0$), using the fact that the poles are continuous functions of these parameters, the poles must always be in the upper half-plane. If, for γ_1 or $\gamma_2 \neq 0$, the imaginary part of the pole were zero, for this the corresponding mode of vibration there would be no damping. So we would have coupled two dissipative systems, and from their combination a conservative system, which is highly improbable.

Then for all values of $\lambda_{12}, \lambda_{21} \neq 0$, we get $|\omega_+^{(1)}| = |\omega_-^{(1)}| < |\omega_+^{(2)}| = |\omega_-^{(2)}|$. The Green's function for the general case is:

$$\begin{aligned} G_0(t,t') = & -i\Theta(t'-t) \sum_1^2 {}_h \frac{e^{i\omega^{(h)}(t-t')}}{(\omega_-^{(h)} - \omega_+^{(h)})(\omega_-^{(h)} - \omega_+^{(k)})(\omega_-^{(h)} - \omega_-^{(k)})} \\ & \cdot \begin{pmatrix} -(\omega_-^{(h)})^2 + 2\gamma_2\omega_-^{(h)} + \omega_2^2 & \lambda_{12} \\ \lambda_{21} & -(\omega_-^{(h)})^2 + i\gamma_1\omega_-^{(h)} + \omega_1^2 \end{pmatrix} \\ & + i\Theta(t-t') \sum_1^2 {}_h \frac{e^{i\omega_+^{(h)}(t-t')}}{(\omega_+^{(h)} - \omega_-^{(h)})(\omega_+^{(h)} - \omega_-^{(k)})(\omega_+^{(h)} - \omega_+^{(k)})} \\ & \cdot \begin{pmatrix} -(\omega_+^{(h)})^2 + i\gamma_2\omega_+^{(h)} + \omega_2^2 & \lambda_{12} \\ \lambda_{21} & -(\omega_+^{(h)})^2 + i\gamma_1\omega_+^{(h)} + \omega_1^2 \end{pmatrix}, \end{aligned} \tag{A9}$$

where $h = 1,2$; $k = 1,2$, and $h \neq k$.

B. Determination of $G_0^S(t,t')$ from $G_0(t,t')$

In an obvious notation it is possible to write (A9) as

$$\begin{aligned} G_0(t,t') = & \Theta(t-t')\left[e^{i\omega_+^{(1)}(t-t')}M_1^+ + e^{i\omega_+^{(2)}(t-t')}M_2^+\right] \\ & + \Theta(t'-t)\left[e^{i\omega_-^{(1)}(t-t')}M_1^- + e^{i\omega_-^{(2)}(t-t')}M_2^-\right]. \end{aligned} \tag{A10}$$

From Eq. (15b) we have:

$$\begin{aligned} G_0^S(t,t') = & \sum_{-\infty}^{\infty} G_0(t,t'+nT)e^{in\omega T} \\ = & G_0(t,t') + e^{i\omega_+^{(1)}(t-t')}M_1^+\left(\sum_1^\infty {}_n e^{inT(\omega-\omega_+^{(1)})}\right) \\ & + e^{i\omega_+^{(2)}(t-t')}M_2^+\left(\sum_1^\infty {}_n e^{inT(\omega-\omega_+^{(2)})}\right) \\ & + e^{i\omega_-^{(1)}(t-t')}M_1^-\left(\sum_1^\infty {}_n e^{-inT(\omega-\omega_-^{(1)})}\right) \\ & + e^{i\omega_-^{(2)}(t-t')}M_2^-\left(\sum_1^\infty {}_n e^{-inT(\omega-\omega_-^{(2)})}\right), \end{aligned}$$

with the definition $g(z) = \Sigma_1^\infty e^{nTz} = (e^{-Tz} - 1)^{-1}$ we find (17).

APPENDIX B: DERIVATION OF THE LINEARIZED FUSION MODEL AT THE CRITICAL SURFACE

We consider the two component (electrons and ions) fluid model and summarize the friction with a constant parameter Γ_μ ($\mu \equiv i,e$). The basic plasma equations are:

$$\left(\frac{\partial}{\partial t} + 2\Gamma_\mu + \mathbf{V}_\mu\nabla\right)\mathbf{V}_\mu = \frac{e_\mu}{m_\mu}\left(\mathbf{E} + \frac{\mathbf{V}_\mu}{e} \times \mathbf{H}\right) - \frac{1}{m_\mu N_\mu}\nabla p_\mu$$

and

$$\frac{\partial N_\mu}{\partial t} = -\nabla(N_\mu \mathbf{V}_\mu), \tag{B1}$$

e_μ, m_μ are the particle charge and mass respectively, N_μ the number density, and p_μ the hydrodynamic pressure. The fields **E**, **H**, which also contain the external laser field, satisfy the Maxwell equations (in Gaussian units):

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t},$$

$$\nabla\times\mathbf{H}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi}{c}\mathbf{J},$$

$$\nabla\cdot\mathbf{H}=0, \tag{B2}$$

$$\nabla\cdot\mathbf{E}=4\pi\Sigma_\mu e_\mu N_\mu,\quad \mathbf{J}=\Sigma_\mu e_\mu N_\mu\mathbf{V}_\mu.$$

The linearization procedure is based on the strength of the laser field $\mathbf{E}_0=\mathbf{E}_0\cos(k_Lz-\omega_Lt)$ (plane wave propagating in the z direction with frequency ω_L):

$$N_\mu=n_0+n_\mu,\quad \mathbf{V}_\mu=\mathbf{V}_{0_\mu}+\mathbf{V}_\mu^{(1)},\quad \mathbf{E}_0=\mathbf{E}_0+\mathbf{E}_1, \tag{B3}$$

$$\mathbf{H}=\mathbf{H}_0+\mathbf{H}_1,\quad P_\mu=p_\mu^{(0)}+p_\mu.$$

The equations simplify considerably because of the following physically valid arguments:

(a) the external field $\mathbf{E}_0$ may be treated as spatially homogeneous (dipole approximation) in the critical density surface defined by $\omega_e\equiv\omega_L$;

(b) the motion of the ions in the direction of the field is neglected ($m_i\gg m_e$ and $\mathbf{V}_0\approx0$).

By eliminating the $\mathbf{E}$ field from the linearized equations we easily obtain:

$$\left(\frac{\partial}{\partial t}+2\Gamma_e+\mathbf{V}_0\cdot\nabla\right)\left(\frac{\partial}{\partial t}+\mathbf{V}_0\cdot\nabla\right)n_e$$
$$=\omega_e^2(n_i-n_e)+S_e^2\nabla^2n_e, \tag{B4}$$
$$\left(\frac{\partial^2}{\partial t^2}+2\Gamma_i\frac{\partial}{\partial t}+\omega_i^2-S_i^2\nabla^2\right)n_i=\omega_i^2n_e,$$

and from the Fourier integrals:

$$n_e=(2\pi)^{-3/2}\int d^3\mathbf{k}\,\tilde{n}_e(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}},$$
$$n_i=(2\pi)^{-3/2}\int d^3\mathbf{k}\,\tilde{n}_i(\mathbf{k},t)e^{i\mathbf{k}\mathbf{r}}. \tag{B5}$$

if we insert these representations into the equations we finally obtain Eqs. (21).

We remember that the relation between the hydrodynamic pressure p_μ and the density n_μ varies in isothermal or adiabatic conditions.

The laser fusion case is almost isothermal:

$$\nabla p_\mu=m_\mu S_\mu^2\nabla n_\mu,\quad S_\mu^2=\frac{1}{m_\mu}\theta_\mu,$$

where θ_μ is the temperature in energetic units.

APPENDIX C: SPECIFIC KRONIG–PENNEY MODELS

The Nishikawa model becomes exactly solvable if $Z(t)=Z_0\sin\omega_Lt$ is replaced by

$$\tilde{Z}(t)=Z_0\sum_0^\infty{}_n\frac{\sin(2n+1)\omega_Lt}{2n+1}=\mp\frac{\pi}{4}Z_0,$$
$$t\gtrless0,\quad -\frac{T}{2}\leqslant t\leqslant\frac{T}{2}, \tag{C1}$$

$\tilde{Z}(t)$ differing from the original $Z(t)$ owing to the presence of the infinite set of the harmonics $n=1,2,...$, which have weights tending to zero when $n\to\infty$. As an extreme limit of the square well approximation one can take the periodic delta function sequence:

$$\tilde{Z}(t)=Z_0\Gamma\left[\delta\left(t-\frac{T}{4}\right)-\delta\left(t+\frac{T}{4}\right)\right],\quad \Gamma=2/\omega_L. \tag{C2}$$

Subsection IV B shows how (C2) reduces the integral equation (19) to the algebraic one (24), from which the determinantal equation (26) follows. Identical final results can be obtained with the direct method by imposing the continuity prescription to the first and the third row of the principal matrix:

$$X_i^j\left(\pm\frac{T^+}{4}\right)-X_i^j\left(\pm\frac{T^-}{4}\right)=0,$$
$$i=1,2;\quad j=1,2,3,4, \tag{C3a}$$

and the prescribed jumps to the elements of the second and fourth row at the discontinuity points:

$$\dot{X}_i^j\left(\pm\frac{T^+}{4}\right)-\dot{X}_i^j\left(\pm\frac{T^-}{4}\right)$$
$$=\pm\Gamma Z_0\left[\mu_{ii}X_i^j\left(\pm\frac{T}{4}\right)+\mu_{il}X_l^j\left(\pm\frac{T}{4}\right)\right], \tag{C3b}$$

where $i=1,2;\ l=2,1;\ j=1,2,3,4$.

The same extreme model can be applied to the linearized laser fusion model rewritten in the form

$$\ddot{X}+\gamma_1\dot{X}+\omega_1^2X=\lambda_{12}Y+\alpha(t)\dot{X}+\beta(t)X, \tag{C4a}$$
$$\ddot{Y}+\gamma_2\dot{Y}+\omega_2^2Y=\lambda_{21}X, \tag{C4b}$$

where

$$\mathbf{V}_0=\tilde{\mathbf{V}}_0\sin\omega_Lt, \tag{C4c}$$
$$X=\tilde{n}_e\quad \omega_1^2=\omega_e^2+k^2S_e^2+\tfrac{1}{2}(\mathbf{k}\cdot\mathbf{V}_0)^2,$$
$$\lambda_{12}=\omega_e^2,\quad \gamma_1=2\Gamma_e, \tag{C4d}$$
$$Y=\tilde{n}_i\quad \omega_2^2=\omega_i^2+k^2S_i^2,$$
$$\lambda_{21}=\omega_i^2,\quad \gamma_2=2\Gamma_i, \tag{C4e}$$
$$\alpha(t)=-2i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\sin\omega_Lt, \tag{C4f}$$

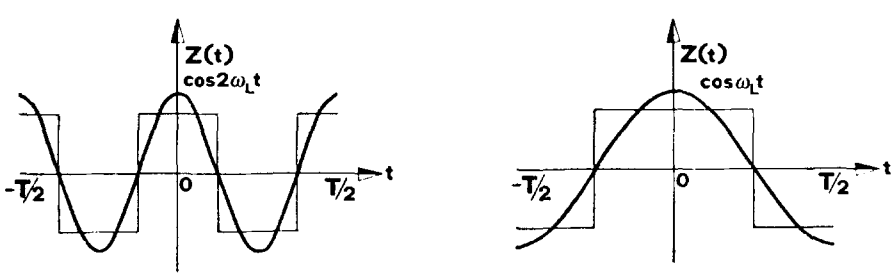

FIG. 4. Periodic square well models for $\cos\omega_Lt$ and $\cos2\omega_Lt$.

$$\beta(t) = -i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\omega_L \cos\omega_L t - 2i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\Gamma_e \sin\omega_L t + \tfrac{1}{2}(\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2 \cos 2\omega_L t. \quad \text{(C4g)}$$

The Kronig–Penney approximation for $Z_1(t) = \sin\omega_L t$ has been considered above for the Nishikawa model. Along this line we sketch in Fig. 4 the approximation introduced in order to replace the functions $Z_2(t) = \cos\omega_L t$ and $Z_3(t) = \cos 2\omega_L t$.

Now if we push forward the periodic square well model up to the extreme delta function configuration, we have

$$Z_1(t) = \Gamma[-\delta(t + T/4) + \delta(t - T/4)], \quad \text{(C5a)}$$

$$Z_2(t) = \Gamma[-\tfrac{1}{2}\delta(t + T/2) + \delta(t) - \tfrac{1}{2}\delta(t - T/2)], \quad \text{(C5b)}$$

$$\tilde{Z}_3(t) = \Gamma[\tfrac{1}{4}\delta(t + T^+/2) - \tfrac{1}{2}\delta(t + T/4) + \tfrac{1}{2}\delta(t) - \tfrac{1}{2}\delta(t - T/4) + \tfrac{1}{4}\delta(t - T^-/2)], \quad \text{(C5c)}$$

where $\Gamma = 2/\omega$.

After some algebraic manipulations:

$$G^S(t,t') = G_0^S(t,t') + S^1(t,-T/2)G^S(-T/2,t') + S^2(t,-T/4)G^S(-T/4,t') + S^3(t,0)G^S(0,t') + S^4(t,T/4)G^S(T/4,t') + S^5(t,T/2)G^S(T/2,t'), \quad \text{(C6)}$$

where

$$S^1\left(t,-\frac{T}{2}\right) = \frac{\Gamma}{4} - 2i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\omega_L + \tfrac{1}{2}(\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2 \times G_0^S\left(t,-\frac{T}{2}\right)\xi_+,$$

$$S^2\left(t,-\frac{T}{4}\right) = \left[-\frac{\Gamma}{2}[-4i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\Gamma_e + \tfrac{1}{2}(\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2]G_0^S\left(t,-\frac{T}{4}\right) - 2i\Gamma\mathbf{k}\cdot\tilde{\mathbf{V}}_0\dot{G}_0^S\left(t,-\frac{T}{4}\right)\right]\xi_+,$$

$$S^3(t,0) = \frac{\Gamma}{2}[2i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\omega_L + \tfrac{1}{2}(\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2]G_0^S(t,0)\xi_+,$$

$$S^4\left(t,\frac{T}{4}\right) = \left[-\frac{\Gamma}{2}[4i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\Gamma_e + \tfrac{1}{2}(\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2]G_0^S\left(t,\frac{T}{4}\right) + 2i\Gamma\mathbf{k}\cdot\tilde{\mathbf{V}}_0\dot{G}_0^S\left(t,\frac{T}{4}\right)\right]\xi_+,$$

$$S^5\left(t,\frac{T}{2}\right) = \frac{\Gamma}{4}[-2i\mathbf{k}\cdot\tilde{\mathbf{V}}_0\omega_L + \tfrac{1}{2}(\mathbf{k}\cdot\tilde{\mathbf{V}}_0)^2]G_0^S\left(t,\frac{T}{2}\right)\xi_+,$$

we have four equations in the form:

$$G_{1j}^S(t,t') = G_{1j}^S(t,t') + \sum_1^5 {}_kS_{11}^k(t,t_k)G_{1j}^S(t_k,t'), \quad \text{(C7a)}$$

$$G_{2j}^S(t,t') = G_{2j}^S(t,t') + \sum_1^5 {}_kS_{21}^k(t,t_k)G_{1j}^S(t_k,t'), \quad \text{(C7b)}$$

where $t_k(-3/4 + k/4)T$. Solving (C7a) for $t = t_{k'} = (-3/4 + k'/4)T$, we have

$$\sum_1^5 {}_k[S_{11}^k(t_{k'},t_k) - \delta_{k,k'}]G_{1j}^S(t_k,t') = -G_{1j}^S(t_{k'},t'), \quad \text{(C8)}$$

where $k' = 1,2,\ldots,5$. Equations (C8) give us two nonsingular 5×5 matrices (one for $j = 1$ and another for $j = 2$).

Therefore, from (C8) we get an explicit form for $G_{1j}^S(t_k,t'), j = 1,2;\ k = 1,2,3,4,5$.

By inserting $G_{1j}^S(t_k,t')$ in (C7a) and (C7b) we find all four matrix elements of $G^S(t,t')$.

The dispersion relation for the laser fusion model in the delta function approach is given by the determinantal equation

$$\det A = 0, \quad \text{(C9a)}$$

where

$$A_{ij} = (S_{11}^i(t_j,t_i) - \delta_{ij}), \quad t_\mu = \left(-\frac{3}{4} + \frac{\mu}{4}\right)T, \quad \text{(C9b)}$$

$\mu \equiv i,j = 1,\ldots,5$.

APPENDIX D: FOURIER ANALYSIS OF PARAMETRIC EXCITATION

In our basic equation (1) the functional dependence of the external excitation term is given by periodic functions (in particular sinusoidal). We can therefore apply the Fourier analysis to the solution $X(t)$.[13] More specifically, from conditions (13) we can write

$$X(t) = e^{i\omega t}Y(t), \quad Y(t+T) = Y(t). \quad \text{(D1)}$$

By introducing into Eq. (1) the Fourier expansions of the periodic functions $Z(t)$, $Y_\mu(t)$ $(\mu = 1,2)$

$$Z(t) = \sum {}_nZ_ne^{in\omega_L t}, \quad X_\mu(t) = \sum {}_nx_{\mu n}e^{i(\omega + n\omega_L)t}, \quad \text{(D2)}$$

one obtains the infinite system

$$\begin{pmatrix} D_1 & -\lambda_{12} \\ -\lambda_{21} & D_2 \end{pmatrix}\begin{pmatrix} x_{1,m} \\ x_{2,m} \end{pmatrix} = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix}\begin{pmatrix} \sum {}_nZ_nx_{1,m-n} \\ \sum {}_nZ_nx_{2,m-n} \end{pmatrix}, \quad \text{(D3)}$$

or

$$\begin{pmatrix} x_{1,m} \\ x_{2,m} \end{pmatrix} = \frac{1}{D}\begin{pmatrix} D_2 & \lambda_{12} \\ \lambda_{21} & D_1 \end{pmatrix}\begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} \times \begin{pmatrix} \sum {}_nZ_nx_{1,m-n} \\ \sum {}_nZ_nx_{2,m-n} \end{pmatrix}, \quad \text{(D4a)}$$

where

$$D_\mu \equiv D_\mu(\omega + n\omega_L) = (\omega + n\omega_L)^2 + i\gamma_\mu(\omega + n\omega_L) + \omega_\mu^2,$$
$$D = D_1D_2 - \lambda_{12}\lambda_{21}. \quad \text{(D4b)}$$

Similar equations are obtained by starting from the Fourier transform $\tilde{X}_\mu(\omega') = (2\pi)^{-1/2}\int_{-\infty}^{+\infty}\exp(-i\omega' t)X_\mu(t)\,dt$ of Eq. (1),

$$\begin{pmatrix} \tilde{X}_1(\omega') \\ \tilde{X}_2(\omega') \end{pmatrix} = \frac{1}{D(\omega')}\begin{pmatrix} D_2(\omega') & \lambda_{12} \\ \lambda_{21} & D_1(\omega') \end{pmatrix}$$

$$\times\begin{pmatrix}\mu_{11} & \mu_{12}\\ \mu_{21} & \mu_{22}\end{pmatrix}\begin{pmatrix}\sum_n Z_n\tilde{X}_1(\omega'-n\omega_L)\\ \sum_n Z_n\tilde{X}_2(\omega'-n\omega_L)\end{pmatrix}. \quad \text{(D5)}$$

If we also take the Fourier transform of condition (13) we have

$$\tilde{X}_\mu(\omega')[e^{i\omega' T}-e^{i\omega T}]=0; \quad \text{(D6)}$$

we obtain the discrete set of harmonics

$$\omega'=\omega+n\omega_L \quad \text{(D7)}$$

(coinciding with the series spectrum). From (D7) the identification of (D5) with Eqs. (D4) immediately follows.

[1]K. Nishikawa, J. Phys. Soc. Japan **24**, 916, 1152 (1968).

[2]K.A. Brueckner and S. Jorna, Rev. Mod. Phys. **46**, 325 (1974).

[3]Martin V. Goldman, Ann. Phys. **38**, 95, 117 (1966); D.F. DuBois and B. Bezzerides, Phys. Rev. A **14**, 1869 (1976). See also the reviews: R.N. Franklin, Rep. Progr. Phys. **40**, 1369 (1977); F.F. Cop, *Handbook on Plasma Instabilities* (Academic, New York, 1978), Vol. 2.

[4]For a review on the Padé method see: G.A. Baker and J. Gammel, *The Padé Approproximant in Theoretical Physics* (Academic, New York, 1970); P.R. Graves-Morris, *Padé Approximants* (Inst. Phys., London, 1972); *Padé Approximants and Their Applications* (Academic, New York, 1973).

[5]A. Sommerfeld, *Lectures on Theortical Physics* (Academic, New York, 1954), Vol. IV, p. 199.

[6]F.W. Byron and R.W. Fuller, *Mathematics of Classical and Quantum Physics* (Addison-Wesley, Reading, Massachusetts, 1970).

[7]For Floquet's theory of periodic systems see: A.D. Myskis, *Advanced Mathematics* (Mir., Moscow, 1975).

[8]In solid state physics, where periodic lattice potentials occur, Eqs. (14) are known as Bloch boundary conditions, while the corresponding Floquet solutions are also called Block functions. The difference between the approach of classical mechanics, adopted in this paper, and the usual quantum mechanical description of solid state systems lies in the fact that we are concerned with the equation of motion and consequently the time variable whereas in quantum mechanics one deals with the stationary state of the potential and periodicity is on the space variables. The differential operator is not necessarily hermitian in our case. The formal correspondence that we established between the oscillator coupled equations and the quantum mechanical formalism of a particle in a lattice guarantees the consistency of the two approaches: the one defined on the whole real axis (associated with the asymptotic Sommerfeld behavior) and the other defined on a period (associated with Bloch conditions). In fact, the propagation of a wave through a lattice can be thought of as a coherent multiple scattering by the various lattice elements. This point of view is developed in J.B. Pendry, *Low Energy Electron Diffraction* (Academic, New York, 1974), Sec. IIIB.

[9]The interpretation of ω_k as a cyclic frequency becomes much more evident in the context of the Fourier analysis of the amplitudes (see Appendix D).

[10]For more details on SGF see the pioneer article: W. Kohn and N. Rostoker, Phys. Rev. **94**, 1111 (1954), and the review: T. Lukes, *Solid State Theory*, edited by P.I. Landsberg (Wiley, New York, 1960), part F, keeping in mind that these concern the quantum mechanical case of a particle in a lattice.

[11]For the resolvent SGF see: T. Lukes and M. Roberts, Proc. Phys. Soc. **91**, 211 (1967), and also T.A. Minelli, Nuovo Cimento B **33**, 619 (1976), where the statements of this subsection are justified.

[12]The process of substituting $\tilde{Z}(t),\tilde{\alpha},\tilde{\beta}$ to $Z(t),\alpha,\beta$ is equivalent, mathematically, to adding an infinite set of higher harmonics (the Fourier expansion terms) to the fundamental frequency ω_L, with a consequent light perturbation of the initial system under examination. The addition of higher harmonics is legitimate as long as one can control for the fact that their weights are rapidly tending to zero (see also the Introduction on this point). The procedure has the advantage of allowing an exact evaluation of the equations without altering qualitatively and, to a reasonable extent, quantitatively, the physical system. (see also in Appendix C the delta function approximation.) For an introduction to the Kronig–Penney models see: L. Brillouin, *Wave Propagation in Periodic Structure* (Dover, New York, 1946), Chap. 8, and also R.A. Smith, *Wave Mechanics of Crystalline Solids* (Chapman, London, 1961).

[13]The Fourier analysis of differential equations with periodic coefficients has been developed by Hill. See, for instance: E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis* (Cambridge, Univ. Press, 1935), Chap. 19.

Normal form for mirror machine Hamiltonians[a)]

Alex J. Dragt and John M. Finn[b)]

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 28 June 1978; revised manuscript received 29 December 1978)

A systematic algorithm is developed for performing canonical transformations on Hamiltonians which govern particle motion in magnetic mirror machines. These transformations are performed in such a way that the new Hamiltonian has a particularly simple *normal form*. From this form it is possible to compute analytic expressions for gyro and bounce frequencies. In addition, it is possible to obtain arbitrarily high order terms in the adiabatic magnetic moment expansion. The algorithm makes use of Lie series, is an extension of Birkhoff's normal form method, and has been explicitly implemented by a digital computer programmed to perform the required algebraic manipulations. Application is made to particle motion in a magnetic dipole field and to a simple mirror system. Bounce frequencies and locations of periodic orbits are obtained and compared with numerical computations. Both mirror systems are shown to be insoluble, i.e., trajectories are not confined to analytic hypersurfaces, there is no analytic third integral of motion, and the adiabatic magnetic moment expansion is divergent. It is expected also that the normal form procedure will prove useful in the study of island structure and separatrices associated with periodic orbits, and should facilitate studies of breakdown of adiabaticity and the onset of "stochastic" behavior.

1. INTRODUCTION AND NOTATION

In the study of a complicated dynamical system, one almost invariably seeks at a minimum to learn the answers to two fundamental questions. First, what areas of phase space are in fact accessible to the system for a given trajectory or class of trajectories? Second, where are the periodic and quasiperiodic orbits, and what are their frequencies? Thus, for example, in the study of magnetic mirror machines one uses the magnetic moment "invariant" to "infer" that certain particles will indeed mirror and will not escape through the ends of the machine. In addition, one develops various expressions or runs numerical codes to determine gyro frequencies, bounce frequencies, and those orbits for which these frequencies are commensurate.

The purpose of this paper is to show how these questions can be studied in detail for mirror machines. Our method makes use of algebraic manipulations performed by a digital computer. We are able to produce analytic expressions for the frequencies and initial conditions associated with periodic and quasiperiodic orbits.[1-3] These expressions should prove to be useful in the study of island structure and separatrices associated with periodic orbits.[4,5] In addition, we are able to obtain arbitrarily high order terms in the complete adiabatic magnetic moment expansion. This latter result has already proved useful in demonstrating the "insolubility" of certain mirror machine problems,[4] and should facilitate studies of the breakdown of adiabaticity and the onset of "stochastic" behavior[6,7] In particular, it has been shown for certain mirror machines that trajectories are not confined to analytic hypersurfaces in phase space. As a result, the adiabatic magnetic moment expansion is divergent, and one can make no mathematically rigorous statement about confinement or the long-term behavior of orbits.[4] Such may in fact be the case for all mirror machines.

More precisely, the purpose of this paper is to show that a certain class of Hamiltonians can be brought *systematically* to a particularly simple "normal form" by a sequence of canonical transformations. The class of Hamiltonians of interest will be called "mirror machine" Hamiltonians since they arise naturally in the study of mirror machines designed for plasma containment. By the word "systematically," we mean there exists an algorithm for analytical computation which can be explicitly implemented by a digital computer programmed to perform certain algebraic manipulations.

The meaning of the term "normal form" will be delineated further after the introduction of suitable mathematical machinery. For the moment, we make the following analogy: In the study of a linear operator or matrix, it is often useful to perform similarity transformations to bring the matrix to diagonal or Jordan canonical form. Once this is done, it is a simple matter to read off the eigenvalues and eigenvectors, to evaluate functions of the matrix such as its exponential and inverse, and to find matrices which will commute with the given matrix. In the study of a classical mechanics problem specified by a certain Hamiltonian, one can try to proceed in a similar spirit. One performs canonical transformations on the Hamiltonian in the hope of bringing it to a

[a)]Supported in part by the National Science Foundation under Grant No. GP-41822X.

[b)]Present address: Naval Research Laboratory, Washington, D.C. 20375

 0022-2488/79/122649-12$01.00

simpler form. Exactly what simple forms a given Hamiltonian can be brought to is not as yet completely known and is still an area requiring active study. (Canonical transformations are in general nonlinear, and thus the problem is intrinsically far more complicated.) However, we will show that there is a normal form for any mirror machine Hamiltonian from which it is possible to compute the frequencies and initial conditions associated with periodic and quasiperiodic orbits. Thus, with the normal form method, it is possible to compute analytic expressions for bounce frequencies and for closed orbits. In addition, the normal form we will describe makes possible the construction of formal integrals of motion for the Hamiltonian in question. Integrals of motion are functions of phase space variables which do not explicitly involve the time and which remain constant on trajectories. In the case of mirror machines, the integral of motion proves to be the *complete adiabatic magnetic moment expansion.* Consequently, it is possible to obtain arbitrarily high order terms in the complete adiabatic magnetic moment expansion providing one is willing to spend sufficient computer time. Finally, the normal form method for mirror machines is related to similar transformation methods which have recently proven to be very useful in such diverse areas as celestial mechanics and molecular physics for both deep mathematical proofs and practical calculations.[8] Thus we are at the threshold of a unified treatment of a wide variety of classical mechanics problems.

Since our work requires the execution of a long sequence of canonical transformations and also the inversion of these transformations, the remainder of this section is devoted to the development of notation and a brief review of the method of Lie transformations which we have found to be particularly useful. In Sec. 2 we specify the nature of a mirror machine Hamiltonian and develop the normal form algorithm. Section 3 shows how use of the normal form algorithm leads to the construction of integrals of motion. Section 4 illustrates the application of the normal form method to two examples of charged particle motion in magnetic mirror fields, namely the magnetic dipole field and that of a simple model mirror machine. Comparisons are made for these two problems between numerical and analytical results. In particular, we study the frequencies of periodic orbits, the constancy of the series for the complete adiabatic magnetic moment expansion, and the insolubility of the simple model mirror machine. A final section summarizes the conclusions of this paper.

The general problem of interest will have n degrees of freedom described by the canonical coordinates $q_1,q_2,\ldots,q_n$ and $p_1,p_2,\ldots,p_n$. For compactness of notation we have found it convenient to treat the q's and p's together by introducing the $2n$ variables $z_1,\ldots,z_{2n}$ defined by the relations

$$z_i = q_i; \quad z_{n+i} = p_i, \quad i = 1 \text{ to } n. \tag{1.1}$$

The method of Lie transformations makes essential use of Poisson brackets and the Lie algebraic structure associated with them. We shall briefly review here the tools needed for this paper. A more detailed explication with proofs has been given earlier.[9]

Suppose $f(z)$ is a particular function defined on phase space. We associate with f the *Lie operator* F by the rule that if g is any other function, then F acting on g is defined by

$$Fg = [f,g]. \tag{1.2}$$

Here the bracket [,] denotes the Poisson bracket. Note that F is linear.

Next we define the linear operator $\exp(F)$, called the *Lie transformation* associated with F and f, by the rule

$$\exp(F) = \sum_0^\infty F^j/j! \tag{1.3}$$

with the convention $F^0 = I$.

Lie transformations have several remarkable properties. Suppose d and e are any two functions. Then we find

$$\exp(F)(de) = (\exp(F)d)(\exp(F)e) \tag{1.4}$$

and

$$\exp(F)[d,e] = [\exp(F)d,\exp(F)e]. \tag{1.5}$$

Consequently, if we define new variables $\bar{z}_i$ by the rule

$$\bar{z}_i(z) = \exp(F)z_i, \tag{1.6}$$

then we have

$$[\bar{z}_i,\bar{z}_j] = [\exp(F)z_i,\exp(F)z_j]$$

$$= \exp(F)[z_i,z_j] = [z_i,z_j]. \tag{1.7}$$

Here we have used (1.5) and the fact that $[z_i,z_j]$ is a number and hence is unchanged by $\exp(F)$. It follows from (1.7) that the new variables $\bar{z}(z)$ are related to the old variables z by a canonical transformation.

Conversely, if the $\bar{z}(z)$ are new variables related to the old variables z by a *canonical transformation near the identity* of the form

$$\bar{z}_i(z) = z_i + \text{higher-degree terms}, \tag{1.8}$$

then it can be shown that there exists a sequence of homogeneous polynomials f_3, f_4, etc., of degree 3, 4, etc., such that

$$\bar{z}_i(z) = \cdots\exp(F_5)\exp(F_4)\exp(F_3)z_i. \tag{1.9}$$

Similarly, the inverse to the transformation (1.8) or (1.9) can be written as

$$z_i(\bar{z}) = \exp(-F_3)\exp(-F_4)\exp(-F_5)\cdots\bar{z}_i. \tag{1.10}$$

In this latter expression the f's are considered as depending on the variables $\bar{z}$, i.e., $f_3(\bar{z}), f_4(\bar{z})$, etc.; and all Poisson brackets are taken with respect to the variables $\bar{z}$.

Finally, suppose $\bar{g}(z)$ is a function defined in terms of some other function $g(z)$ by the rule

$$\bar{g}(z) = g(\bar{z}(z)), \tag{1.11}$$

where z and $\bar{z}$ are related by (1.9). Then it follows from consideration of a series expansion of g and repeated use of (1.4) that

$$\bar{g}(z) = \cdots\exp(F_5)\exp(F_4)\exp(F_3)g. \tag{1.12}$$

Note that in making canonical transformations, we take the

active viewpoint: trajectories and functions are transformed; the coordinate system remains fixed.

With this preparatory background, we are now able to state more precisely our purpose. Suppose we wish to study the nature of the trajectories governed by a certain mirror machine Hamiltonian. For clarity, we denote this Hamiltonian by the symbol h^{old}. We assume that h^{old} does not depend explicitly on time. Then our aim is to find a sequence of homogeneous polynomials f_3, f_4, etc., such that the transformed or "new" Hamiltonian h^{new}, given by

$$h^{\text{new}} = \cdots \exp(F_5)\exp(F_4)\exp(F_3)\, h^{\text{old}}, \tag{1.13}$$

has a particularly simple form. By "simple," we mean that h^{new} should only depend on certain combinations of the variables z in such a way that it is easy to find functions i^{new}, called integrals of h^{new}, which do not depend explicitly on time and which satisfy the relation

$$[i^{\text{new}}, h^{\text{new}}] = 0. \tag{1.14}$$

Whenever such an i^{new} can be found, then it is immediately possible to find an associated integral i^{old} of the original Hamiltonian. In analogy to (1.13 we define i^{old} in terms of i^{new} by the rule

$$i^{\text{old}} = \exp(-F_3)\exp(-F_4)\exp(-F_5)\cdots i^{\text{new}}. \tag{1.15}$$

We then find, using the definitions and (1.5), that

$$\begin{aligned}[i^{\text{old}}, h^{\text{old}}] &= [\exp(-F_3)\exp(-F_4)\cdots i^{\text{new}}, \exp(-F_3) \\ &\quad \times \exp(-F_4)\cdots h^{\text{new}}] \\ &= \exp(-F_3)\exp(-F_4)\cdots [i^{\text{new}}, h^{\text{new}}] = 0.\end{aligned} \tag{1.16}$$

Hence, i^{old} is an integral of motion for the Hamiltonian h^{old}. That is,

$$\frac{di^{\text{old}}}{dt} = [i^{\text{old}}, h^{\text{old}}] = 0. \tag{1.17}$$

Of course, even when i^{new} is a simple expression in terms of the p's and q's as will prove to be the case in Sec. 3, i^{old} will in general be very complicated because of the Lie transformations indicated in (1.15).

Still more will prove to be possible. It is evident from (1.13), with the aid of (1.11) and (1.12), that h^{old} and h^{new} are related by a canonical transformation. We have $h^{\text{new}}(z) = h^{\text{old}}(\bar{z}(z))$. Therefore, if because of its simple form one can find the frequencies and initial conditions for the periodic and quasiperiodic orbits generated by h^{new}, then it is easy to deduce the equivalent information for h^{old}. We will see in Sec. 4 that this is indeed the case.

2. NORMAL FORM ALGORITHM

We begin by assuming that the canonical coordinates z have been selected in such a way that the origin in phase space is an equilibrium point. Thus if the Hamiltonian h^{old} is expanded about the origin, we obtain an expression of the form

$$h^{\text{old}} = \sum_2^\infty h_i^{\text{old}}(z), \tag{2.1}$$

where each h_i^{old} is a homogeneous polynomial of degree i. Next, we assume that the linearized equations of motion about the equilibrium point have m *zero* frequencies ($m < n$) and $n - m$ *nonzero* frequencies. That is, we assume that with a suitable choice of coordinates h_2^{old} has the form

$$\begin{aligned}h_2^{\text{old}} &= \tfrac{1}{2}(p_1^2 + \cdots + p_m^2) + (\tfrac{1}{2}\alpha_{m+1})(p_{m+1}^2 + q_{m+1}^2) \\ &\quad + \cdots + (\tfrac{1}{2}\alpha_n)(p_n^2 + q_n^2),\end{aligned} \tag{2.2}$$

where all the α's are positive.

Hamiltonians of this form arise naturally in the study of mirror machines. How this comes about in detail will become apparent in Sec. 4, where we study two explicit examples. Roughly speaking, one can say that a degree of freedom for which a frequency is zero corresponds to motion along a magnetic field line. For this motion there is no restoring force in lowest approximation. By contrast, the degrees of freedom associated with nonzero frequencies correspond to motion across field lines; and in this case there is a restoring force even in first approximation.

As explained in the introduction, our goal is to find functions f_3, f_4, etc., such that h^{new} given by (1.13) has a simple form. To study systematically what possibilities exist, it is convenient to introduce the notation

$$h^k = \exp(F_k)\exp(F_{k-1})\cdots\exp(F_3)h^{\text{old}}. \tag{2.3}$$

Then we have, for example, the relations

$$h^3 = \exp(F_3)h^{\text{old}}, \tag{2.4a}$$

$$h^\infty = h^{\text{new}}, \tag{2.4b}$$

and the recursion formula

$$h^k = \exp(F_k)\, h^{k-1}, \quad k \geqslant 3, \quad \text{and} \quad h^2 = h^{\text{old}}. \tag{2.5}$$

In analogy to the notation of (2.1), let us write

$$h^k = \sum_{j=2}^\infty h_j^k \tag{2.6}$$

where each term h_j^k is a homogeneous polynomial of degree j. Then from (2.5) we have the relation

$$\sum_{j=2}^\infty h_j^k = \exp(F_k)\sum_{l=2}^\infty h_l^{k-1}, \quad k \geqslant 3. \tag{2.7}$$

Evidently, Eq. (2.7) implies the equality of terms of like degree. Our problem is to identify them. Let $\mathscr{P}_k$ denote the set of homogeneous polynomials of degree k, and suppose f_i and g_j are two homogeneous polynomials of degree i and j, respectively. Then, since the Poisson bracket operation involves multiplication and two differentiations, we have the relation

$$[f_i, g_j] \in \mathscr{P}_{i+j-2}. \tag{2.8}$$

Employing this observation and the definition (1.3), we find from (2.7) the relations

$$h_2^k = h_2^{k-1} = h_2^{\text{old}}, \tag{2.9a}$$

$$h_j^k = h_j^{k-1}, \quad j < k \quad \text{and} \quad k \geqslant 3, \tag{2.9b}$$

$$h_k^k = h_k^{k-1} + F_k\, h_2^{\text{old}}. \tag{2.9c}$$

We can now draw several conclusions. First, we see that $h_2^{\text{new}} = h_2^{\text{old}}$ and $h_k^{\text{new}} = h_k^k$. Second, the term h_k^{new} depends in a rather complicated way on F_k, $F_{k-1},\cdots,F_3$ and h_k^{old}, $h_{k-1}^{\text{old}},\ldots,h_2^{\text{old}}$. Finally, if our goal is to make h_k^{new} "simple," then by (2.9c), our last chance to do so occurs at the stage at which F_k is determined.

It is apparent that (2.9c) is a key relation. In keeping with the notation of Sec. 1, let H_2 be the Lie operator associated with h_2^{old}. Then we have the relation

$$F_k h_2^{\text{old}} = [f_k, h_2^{\text{old}}] = -H_2 f_k. \tag{2.10}$$

Thus, we can also write (2.9c) in the form

$$h_k^k = h_k^{k-1} - H_2 f_k. \tag{2.11}$$

For further discussion, it is useful to regard all polynomials of degree k as elements of a vector space. Then H_2 may be regarded as a linear operator mapping $\mathscr{P}_k$ on to itself. Evidently, the term $H_2 f_k$ consists of all homogeneous polynomials in $\mathscr{P}_k$ that are in the *range* of the operator H_2. (A vector y is in the range of an operator A if there exists a vector x such that $y = Ax$.) Thus, with the aid of (2.11) we are able, by a suitable choice of f_k, to adjust h_k^{new} by any polynomial in $\mathscr{P}_k$ lying within the range of H_2. This is the fundamental result which we shall use in the rest of this paper.

The exploration of the range of a linear operator is facilitated by the introduction of a scalar product. When a scalar product is defined, the Hermitian adjoint of H_2, denoted by $H_2^\dagger$, is also defined. Indicating the scalar product operation by angular brackets, we have the relation

$$\langle a, H_2^\dagger b\rangle = \langle H_2 a, b\rangle. \tag{2.12}$$

The virtue of the introduction of a scalar product is that we can then use the result that each subspace $\mathscr{P}_k$ can be decomposed into a *direct* sum in the form

$$\mathscr{P}_k = \mathscr{R}_k \oplus \mathscr{N}_k, \tag{2.13}$$

where $\mathscr{R}_k$ denotes the range of H_2 and $\mathscr{N}_k$ denotes the null space of $H_2^\dagger$.

The correctness of this result is easily verified for any operator A. First, note that the range of an operator A is itself a linear vector space. For suppose that y and y' are contained in the range of A. Then there exist vectors x,x' such that $y = Ax$ and $y' = Ax'$. Let α and α' be any two scalars. We have $\alpha y + \alpha' y' = A(\alpha x + \alpha x')$, and hence $\alpha y + \alpha' y'$ is also in the range of A. Now let the vectors $u_1,u_2,\cdots$, form a basis for the range of A. Without loss of generality, they can be selected to be orthonormal thanks to the Gram–Schmidt process.[10] Second, let $v_1,v_2,\cdots$ be the remaining orthonormal basis vectors needed to span the complete space. By construction, the v's can be taken to be orthogonal to the u's, and hence to the range of A. They will then also be in the null space $A^\dagger$. That is, we will have $A^\dagger v_j = 0$. For let w be any vector. We find $\langle w, A^\dagger v_j\rangle = \langle Aw, v_j\rangle = 0$ because Aw is in the range of A. It follows that $A^\dagger v_j = 0$ since w is an arbitrary vector. Conversely, any vector z in the null space of $A^\dagger$ will be orthogonal to the range of A. For if y is in the range of A, we have $\langle y,z\rangle = \langle Ax,z\rangle = \langle x,A^\dagger z\rangle = 0$. The verification is now complete, because any vector can be written as a linear combination of the u's and v's since together they form a basis for the entire space. The portion of the expansion which involves the u's will be in the range of A, and the remaining portion involving the v's will be in the null space of $A^\dagger$.

We next consider the choice of a suitable scalar product. Its discovery requires a bit of trial and error. We have found the following definition to be convenient. Let $|j;m\rangle$ denote the monomial defined by

$$|j_1\cdots j_n;m_1\cdots m_n\rangle = \prod_{i=1}^{n}\left[\frac{(2j_i)!}{(j_i-m_i)!(j_i+m_i)!}\right]^{1/2} p_i^{j_i-m_i} q_i^{j_i+m_i}. \tag{2.14}$$

In this expression each j_i is positive or zero, each j_i satisfies $-j_i \leqslant m_i \leqslant j_i$, and each j_i and m_i is integral or half integral. For $|j;m\rangle$ to belong to $\mathscr{P}_k$, we require $2(j_1+\cdots+j_n) = k$. The monomials $|j;m\rangle$ are linearly independent and clearly form a basis. Our scalar product will be defined by the requirement that they form an orthonormal basis,

$$\langle j';m'|j;m\rangle = \delta_{j'j}\delta_{m'm}. \tag{2.15}$$

Here, the quantity $\delta_{j'j}$ equals $+1$ if all the indices denoted by j' and j are respectively equal, and it is zero otherwise.

The computation of $H_2^\dagger$ is a simple task. From (2.2) we see that h_2^{old} consist of the squares p_i^2 and q_i^2. For these functions we use the notation $\text{ad}(p_i^2)$ and $\text{ad}(q_i^2)$ to denote the associated Lie operators since in this case the capital letter convention is not convenient.[11] Then using (1.2) and (2.14), we find upon computing the required Poisson bracket that

$$\text{ad}(q_i^2)|j_1\cdots j_n;m_1\cdots m_n\rangle = 2(j_i+m_i+1)^{1/2}(j_i-m_i)^{1/2} \times |j_1\cdots j_n;m_1\cdots,m_i+1,\cdots m_n\rangle, \tag{2.16a}$$

$$\text{ad}(p_i^2)|j_1\cdots j_n;m_1\cdots m_n\rangle = -2(j_i-m_i+1)^{1/2}(j_i+m_i)^{1/2} \times |j_1\cdots j_n;m_1\cdots,m_i-1,\cdots m_n\rangle. \tag{2.16b}$$

We observe that, in analogy to quantum mechanics, $\text{ad}(q_i^2)$ behaves like twice an angular momentum raising operator. Similarly, $\text{ad}(p_i^2)$ behaves like twice the negative of an angular momentum lowering operator.[12] It follows immediately or by direct computation that

$$\text{ad}(q_i^2)^\dagger = -\text{ad}(p_i^2), \tag{2.17a}$$

$$\text{ad}(p_i^2)^\dagger = -\text{ad}(q_i^2). \tag{2.17b}$$

We are ready to specify our choice of f_k in relation (2.11). Using (2.13), we can uniquely write

$$h_k^{k-1} = r_k + n_k, \tag{2.18}$$

where r_k is in the range of H_2 and n_k is in the null space of $H_2^\dagger$. Next, we require that f_k satisfy the equation

$$H_2 f_k = r_k. \tag{2.19}$$

This equation always has a solution because r_k is in the range

of H_2 by construction.[13] With this choice for f_k, we find from (2.11) and (2.18) the result

$$h_k^{\text{new}} = h_k^k = n_k, \quad k \geqslant 3. \tag{2.20}$$

That is, it is always possible to choose f_k in such a way that h_k^{new} for $k \geqslant 3$ is in the null space of $H_2^\dagger$.

At this point we should make clear to the reader that we do not maintain that the imposition of (2.19) is always the optimal procedure. Indeed, we are studying other strategies, which will be the subject of another paper.[14] However, the above procedure is clearly a mathematically attractive option worth exploring. We shall see in the next section that it has interesting physical consequences because it leads directly to the formal construction of integrals of motion.

3. INTEGRALS OF MOTION

In the last section, we made a partial exploration of how the choice of the f_k affected the form of h^{new}. We found that by imposing (2.19), it was possible to arrange that each term h_k^{new} in h^{new} (save for h_2^{new}) would be in the null space of $H_2^\dagger$. In this section we will show that this choice leads to the determination of at least one and perhaps several integrals of motion for h^{new}.

Let us express the function h_2^{old} given by (2.2) in the form

$$h_2^{\text{old}} = c + d, \tag{3.1}$$

where c and d are given by

$$c = \tfrac{1}{2}\alpha_{m+1}(p_{m+1}^2 + q_{m+1}^2) + \cdots + \tfrac{1}{2}\alpha_n(p_n^2 + q_n^2), \tag{3.2a}$$

$$d = \tfrac{1}{2}(p_1^2 + \cdots + p_m^2). \tag{3.2b}$$

We shall show that c is an integral of motion for h^{new}. That is, c satisfies the equation $[c, h^{\text{new}}] = 0$.

The proof requires a series of steps. First, suppose that $|n_s\rangle$ is a homogeneous polynomial of degree s. Imagine that $|n_s\rangle$ is expanded as a linear combination of the basis vectors $|j;m\rangle$ given by (2.14). Evidently we must have $2(j_1 + \cdots + j_n) = s$ for every term in the expansion and hence $2j_i \leqslant s$ for every factor in each $|j;m\rangle$. It follows that

$$[\text{ad}(p_i^2)^\dagger]^{s+1}|n_s\rangle = 0 \quad i = 1,2,\cdots,m, \tag{3.3}$$

since $\text{ad}(p_i^2)^\dagger$ is proportional to a raising operator by (2.16a) and (2.17b).

Let C and D denote the Lie operators associated with c and d, respectively, and consider the quantity $[D^\dagger]^{m(s+1)}|n_s\rangle$. Since all the $\text{ad}(p_i^2)^\dagger$ commute with each other, we can expand this quantity to obtain an expression of the form

$$[D^\dagger]^{m(s+1)}|n_s\rangle = \sum_\rho \beta(\rho)[\text{ad}(p_i^2)^\dagger]^{\rho_1}\cdots[\text{ad}(p_m^2)^\dagger]^{\rho_m}|n_s\rangle, \tag{3.4}$$

where the $\beta(\rho)$ are certain coefficients. We note that in each term the exponents are non-negative and must satisfy $\rho_1 + \cdots + \rho_m = m(s+1)$. It follows that in each term there must be at least one exponent ρ_i such that $\rho_i \geqslant (s+1)$. This implies by (3.3) that

$$[\text{ad}(p_i^2)^\dagger]^{\rho_i}|n_s\rangle = 0. \tag{3.5}$$

Therefore, we must have

$$[D^\dagger]^{m(s+1)}|n_s\rangle = 0. \tag{3.6}$$

From (3.1) we have

$$H_2^\dagger = C^\dagger + D^\dagger, \tag{3.7}$$

Also, from (2.17) and (3.2a), C is antihermitian, that is,

$$C^\dagger = -C. \tag{3.8}$$

Solving (3.7) for $D^\dagger$ and inserting the result into (3.6) gives

$$[C + H_2^\dagger]^{m(s+1)}|n_s\rangle = 0. \tag{3.9}$$

Since C and $H_2^\dagger$ commute, the left-hand side of (3.9) can be expanded to give

$$[C^{m(s+1)} + m(s+1)\,C^{m(s+1)-1}H_2^\dagger + \cdots + (H_2^\dagger)^{m(s+1)}]|n_s\rangle = 0. \tag{3.10}$$

Let us now add the further hypothesis, as our notation may already have suggested, that $|n_s\rangle$ is in the null space of $H_2^\dagger$,

$$H_2^\dagger|n_s\rangle = 0. \tag{3.11}$$

Then, all the terms on the left-hand side, except for the first, automatically annihilate $|n_s\rangle$, and we conclude

$$C^{m(s+1)}|n_s\rangle = 0. \tag{3.12}$$

We are almost done. Since C is antihermitian and maps $\mathscr{P}_s$ into itself, we know that its eigenvectors in $\mathscr{P}_s$ must form a complete set in $\mathscr{P}_s$. Thus we can write an expansion of the form

$$|n_s\rangle = \sum_\gamma \beta_\gamma|\gamma\rangle, \tag{3.13}$$

where the β_γ are certain coefficients and the polynomials $|\gamma\rangle$ are linearly independent and satisfy eigenvector relations of the form

$$C|\gamma\rangle = \nu_\gamma|\gamma\rangle. \tag{3.14}$$

Insertion of the expansion (3.13) into (3.12) and use of (3.14) gives the result

$$\sum_\gamma \beta_\gamma \nu_\gamma^{m(s+1)}|\gamma\rangle = 0. \tag{3.15}$$

But, since the polynomials are linearly independent, we must then have $\beta_\gamma \nu_\gamma^{m(s+1)} = 0$ for every γ, which in turn implies $\beta_\gamma \nu_\gamma = 0$ for every γ. From this we conclude that

$$C|n_s\rangle = \sum_\gamma \beta_\gamma \nu_\gamma|\gamma\rangle = 0. \tag{3.16}$$

We have shown that if $|n_s\rangle$ is in the null space of $H_2^\dagger$ it must also be in the null space of C.

The result we have been working to prove now follows immediately. Thanks to our normal form algorithm, we have arranged that each h_k^{new} for $k \geqslant 3$ is in the null space of $H_2^\dagger$, and hence

$$Ch_k^{\text{new}} = 0 \quad \text{for } k \geqslant 3. \tag{3.17}$$

Moreover, it is easily checked that

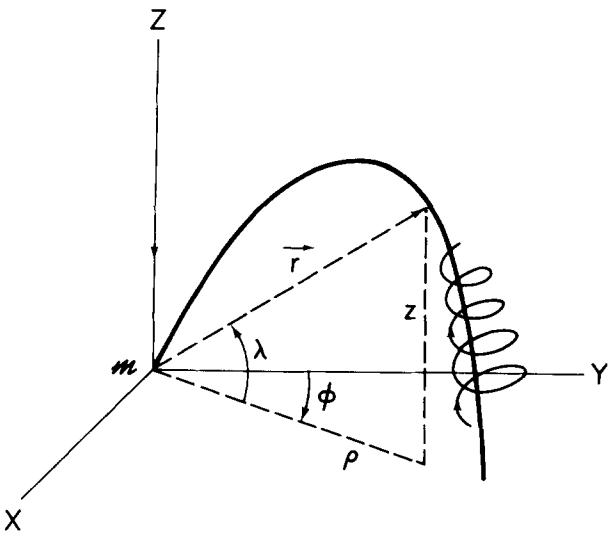

FIG. 1. Motion of a trapped particle in a magnetic dipole field.

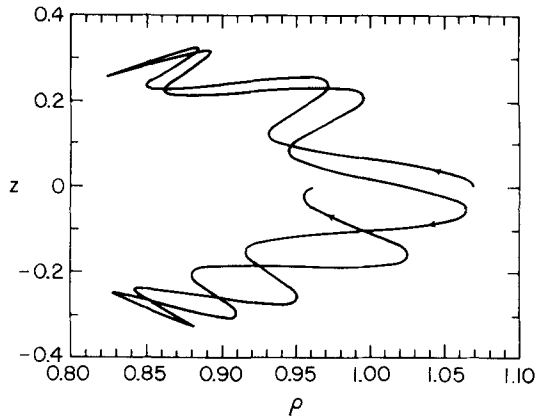

FIG. 2. A trapped orbit as seen in ρ,z coordinates. The initial conditions are $z = 0, \rho = 1.07, \dot{\rho} = 0, \dot{z} = 0.0355$.

$$Ch_2^{\text{new}} = Ch_2^{\text{old}} = [c, h_2^{\text{old}}] = 0. \tag{3.18}$$

Consequently, we have

$$Ch^{\text{new}} = [c, h^{\text{new}}] = 0, \tag{3.19}$$

and c is an integral of motion of h^{new} as advertised.

We have seen that the normal form algorithm of Sec. 2 leads to the determination of an integral of motion i^{new} for h^{new}, namely $i^{\text{new}} = c$. In some cases, for example when the frequencies α_i are irrational in a way as to be incommensurate, it is possible to exhibit additional integrals. This is shown in the Appendix.

It is worth remarking at this point that the normal form algorithm required for mirror machine Hamiltonians is considerably more complicated than that used in celestial mechanics. In the latter case it can be shown that $H_2^\dagger = -H_2$, and then the analysis is far simpler than the preceding has been.

4. EXAMPLES AND APPLICATIONS

In this section we study two Hamiltonian systems which describe the motion of charged particles in magnetic mirror geometries. The first problem considered is that of the motion of a charged particle in a magnetic dipole field, the so-called Stǿrmer problem. This problem is an idealized description of the Van Allen radiation. The second system considered is a simple model mirror machine characterized by the magnetic field given in cylindrical coordinates (ρ,ϕ,z) by

$$\mathbf{B} = (B_0/a^2)[-\rho z\hat{e}_\rho + (a^2 + z^2)\hat{e}_z]. \tag{4.1}$$

The variable a is a typical length scale for the mirror.

Our major tool for dealing with these systems is the use of the normal form algorithm. The algebra involved in carrying out the procedure is very lengthy, but completely routine. Therefore, we have programmed a digital computer to carry out the necessary steps. In brief, we have written routines using the language FORMAL to carry out the decomposition (2.18), solve (2.19) for f_k, and perform (2.5) to move from h^{k-1} to h^k. The calculations were performed through sixth order for the Stǿrmer problem [i.e., $k = 6$ in Eq. (2.3)] and through ninth order for the model mirror machine. Typical calculations required 45 min of Univac 1108 time. We expect that specially written routines for the same purpose which are currently under development will require considerably less computer time.

Figure 1 shows the motion of a typical particle trapped by a magnetic dipole field. When the equations of motion are written in cylindrical coordinates, the axial symmetry associated with a dipole field and scaling of space and time can be used to reduce the problem to the determination of the orbits governed by the reduced Hamiltonian

$$h(p_\rho,p_z,\rho,z) = \tfrac{1}{2}(p_\rho^2 + p_z^2) + \tfrac{1}{2}(1/\rho - \rho/r^3)^2. \tag{4.2}$$

That is, due to axial symmetry, the problem is reduced to one having two degrees of freedom. Once the motion in the ρ,z plane is determined so that $\rho(t)$ and $z(t)$ are known, $\phi(t)$ can be found by a quadrature. Details are given in Refs. 2 and 4. Figure 2 shows a typical orbit as it appears in the ρ,z plane.

The Hamiltonian (4.2) is not in the form of a power series, and consequently we cannot apply the normal form algorithm directly. However, we observe from Figs. 1 and 2

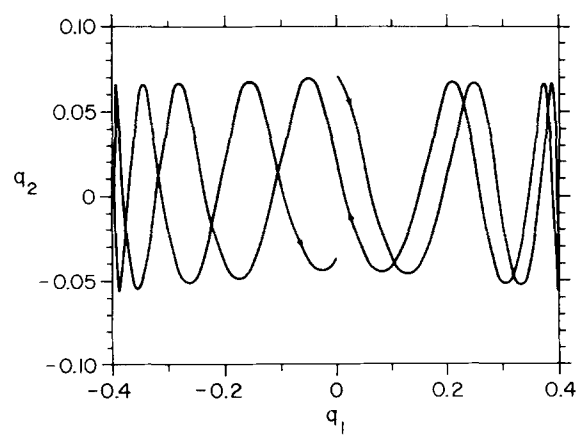

FIG. 3. The orbit of Fig. 2 as it appears in dipolar coordinates.

that the motion consists of gyration about a field line superimposed upon motion along a field line. We therefore introduce orthogonal dipolar coordinates q_1 and q_2 given by

$$q_1 = z/r^3, \quad q_2 = r^3/\rho^2 - 1. \tag{4.3}$$

Roughly speaking, the coordinate q_1 describes motion along the confining field line, i.e., the guiding center motion, and q_2 describes motion perpendicular to the field line. This fact is illustrated in Fig. 3, which displays the orbit of Fig. 2 as it appears in dipolar coordinates. The motion has now been separated, in first approximation, into oscillations about $q_2 = 0$ superimposed upon motion along the q_1 axis.

Now let p_1 and p_2 be momenta canonically conjugate to q_1 and q_2. Then after calculation,[15] one finds that the Hamiltonian (4.2) when expressed in terms of these new variables has a power series expansion. Explicitly, employing the notation (2.1), one finds for the first four terms

$$\begin{aligned} h_2^{\text{old}} &= \tfrac{1}{2}p_1^2 + \tfrac{1}{2}(p_2^2 + q_2^2), \\ h_3^{\text{old}} &= \tfrac{1}{2}(-4q_2^3 - 6q_2p_1^2), \\ h_4^{\text{old}} &= \tfrac{1}{2}(6p_2^2q_1^2 + 21q_2^2p_1^2 + 10q_2^4 + 3q_1^2q_2^2 + 9q_1^2p_1^2). \end{aligned} \tag{4.4}$$

Note that h_2^{old} has the form (2.2).

Before continuing, we should make a remark about the transformation (4.3). It can be shown that for $q_2 = 0$, the transformation from $\lambda = \tan^{-1}(\rho/z)$ to q_1 has a singularity for complex values of q_1 and is analytic only for $|q_1| \leqslant 3(3)^{1/2}/16$.[3] As a consequence, neither the Hamiltonian (4.4) nor the results derived from its normal form are expected to have meaning for $|q_1| \geqslant 3\sqrt{3}/16$.

The stage is set for the application of the normal form algorithm of Sec. 2. The coefficients of the power series expansion (4.4) are inserted into a properly coded computer program. Sometime later the coefficients for the f_k generating (2.3) and the coefficients for the h_k^{new} emerge.

All the results obtained are too lengthy to record here. We find, for example, that f_3 is given by

$$f_3 = -2p_2q_2^2 - 3p_1^2p_2 - \tfrac{4}{3}p_2^3. \tag{4.5}$$

The higher f's rapidly become much longer expressions and are of little direct interest.

The normal form Hamiltonian is of direct interest. Through terms of order 6, and using a notation similar to (3.2a), that is $c_2 = \frac{1}{2}(p_2^2 + q_2^2)$, the normal form Hamiltonian is given by the expression

$$\begin{aligned} h^{\text{new}} = {} & \tfrac{1}{2}p_1^2 + c_2 + (\tfrac{9}{2}c_2q_1^2 - \tfrac{15}{2}c_2^2) \\ & + (\tfrac{39}{8}c_2q_1^4 + \tfrac{9}{16}c_2^2q_1^2 + \tfrac{15}{2}c_2^3). \end{aligned} \tag{4.6}$$

We see that h^{new} is of the functional form

$$h^{\text{new}} = \tfrac{1}{2}p_1^2 + g(c_2, q_1). \tag{4.7}$$

That is, the normal form Hamiltonian is "simple" in the sense that it depends on the variables p_2, q_2 only in the combination $c_2 = \frac{1}{2}(p_2^2 + q_2^2)$. This is, of course, what is to be expected because according to (3.19), c_2 is an integral of motion for h^{new}.

It is also of interest to record some terms of the integral of motion i^{old} obtained from (1.15) with $i^{\text{new}} = c_2$. They are also calculated by our computer program. We write $i^{\text{old}} = i_2^{\text{old}} + i_3^{\text{old}} + i_4^{\text{old}} + \cdots$, etc. The first few terms are given by

$$\begin{aligned} i_2^{\text{old}} &= \tfrac{1}{2}(p_2^2 + q_2^2), \\ i_3^{\text{old}} &= -3p_1^2q_2 - 2q_2^3, \\ i_4^{\text{old}} &= \tfrac{27}{8}p_1^2p_2^2 + \tfrac{117}{8}p_1^2q_2^2 + \tfrac{3}{4}p_2^2q_1^2 + \tfrac{15}{4}p_2^2q_2^2 - \tfrac{3}{4}q_1^2q_2^2 \\ &\quad - \tfrac{3}{2}p_1p_2q_1q_2 + \tfrac{9}{2}p_1^4 + \tfrac{15}{8}p_2^4 + \tfrac{55}{8}q_2^4. \end{aligned} \tag{4.8}$$

The industrious reader is invited to verify for himself that the Poisson bracket (1.17) in fact vanishes using the expansions (4.4) and (4.8). Incidentally, that it does so has been verified directly as a check by a computer programmed Poisson bracket routine.

The series i^{old} has great utility in the examination of the nature of motion in a dipole field in fine detail. It can be used, among other things, to show that the Størmer problem is insoluble. What this means is described extensively elsewhere.[4] We shall give a parallel but much abbreviated treatment of the question of insolubility later on in this section when we discuss the model mirror machine example.

As advertised in the first section of this paper, the normal form Hamiltonian may be sufficiently simple that it is possible to find the frequencies and initial conditions for periodic and quasiperiodic orbits. We shall now see that this is the case for the Størmer problem.

Observe that the Hamiltonian (4.6) is of the form

$$h^{\text{new}} = \beta(c_2) + \tfrac{1}{2}p_1^2 + \tfrac{1}{2}\omega^2(c_2)q_1^2 + \eta(c_2)q_1^4 + \cdots, \tag{4.9}$$

where

$$\begin{aligned} \beta(c_2) &= c_2 - \tfrac{15}{2}c_2^2 + \tfrac{15}{2}c_2^3 + \cdots, \\ \omega^2(c_2) &= 9c_2 + \tfrac{9}{8}c_2^2 + \cdots, \\ \eta(c_2) &= \tfrac{39}{8}c_2 + \cdots. \end{aligned} \tag{4.10}$$

Since c_2 is an integral of motion and therefore constant in time, we conclude that the motion in p_1, q_1 governed by h^{new} is that of an anharmonic oscillator described by the c_2 dependent parameters $\omega^2(c_2)$, $\eta(c_2)$, etc. This circumstance suggests that we should attempt to bring the quadratic part of (4.9) to the form $(\frac{1}{2}\alpha)(p_1^2 + q_1^2)$, which is analogous to (3.2a) as far as the variables p_1, q_1 are concerned, and then we should again apply some normal form algorithm.

Consider the canonical transformation generated by the function g_2 given by

$$g_2 = \tfrac{1}{2}p_1q_1 \log\omega(c_2). \tag{4.11}$$

Note that g_2 is homogeneous of degree 2 as far as the variables q_1, p_1 are concerned. We find that

$$\begin{aligned} \exp(G_2)c_2 &= c_2, \\ \exp(G_2)p_1 &= p_1\omega^{1/2}, \\ \exp(G_2)q_1 &= q_1/\omega^{1/2}. \end{aligned} \tag{4.12}$$

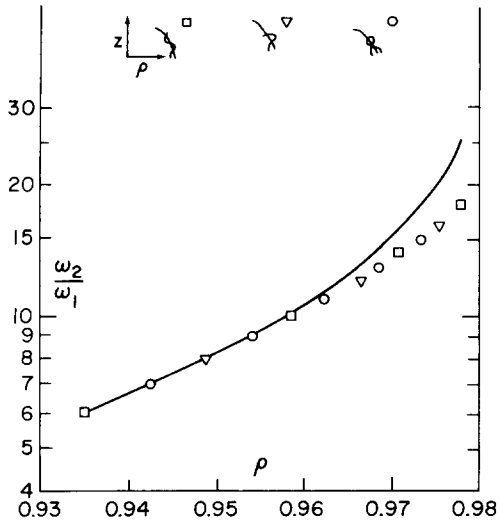

FIG. 4. The frequency ratio ω_2/ω_1, as computed from Eq. (4.17), plotted as a function of the value of ρ. The other initial conditions are $z = 0, \dot{\rho} = 0$, and $h = 0.002907$. For comparison, actual periodic orbits obtained numerically are plotted as points. Initial conditions for these orbits were supplied by R. DeVogelaere (private communication). The type of orbit is shown at the top of the figure.

Consequently, we find that

$$\exp(G_2)h^{\text{new}} = \beta(c_2) + \tfrac{1}{2}\omega(c_2)(p_1^2 + q_1^2) + \gamma(c_2)q_1^4 + \cdots, \tag{4.13}$$

where

$$\gamma(c_2) = \eta/\omega^2. \tag{4.14}$$

We see that apart from the term $\beta(c_2)$, which plays no role in the determination of the p_1, q_1 motion, the transformed Hamiltonian (4.13) has a quadratic part of the form (3.1) with the term of the form (3.2b) completely absent. Now consider the Lie operator associated with this quadratic part. It will be antihermitian because of the analog of (3.8), and hence the null spaces of the analogs of H_2 and $H_2^\dagger$ will coincide in this case.

Let h^* denote the result of applying the normal form algorithm a second time. We write $h^* = \cdots\exp(G_4)\exp(G_3)\exp(G_2)h^{\text{new}}$ where the functions $g_3, g_4, \cdots$ which lead to G_3, $G_4, \cdots$ and which may involve c_2 as a parameter are still to be determined. It follows from the discussion in the previous paragraph that one can arrange to have h^* lie in the null space of $\text{ad}(p_1^2 + q_1^2)$. If this is done, h^* will only depend on the variables p_1, q_1 in the combination $c_1 = \frac{1}{2}(p_1^2 + q_1^2)$. Of course, h^* may also depend on c_2. After calculation employing the normal form algorithm, we find the explicit result

$$h^*(c_1,c_2) = \beta(c_2) + \omega(c_2)c_1 + 13c_1^2/(16 + 2c_2)$$
$$- 25857/62208c_2^{-1/2}[1 + (c_2/8)]^{-5/2}c_1^3 + \cdots. \tag{4.15}$$

Let us introduce variables ϕ_1 and ϕ_2 which are canonically conjugate to the variables c_1 and c_2 by means of the equations

$$q_i = (2c_i)^{1/2}\sin\phi_i, \; p_i = (2c_i)^{1/2}\cos\phi_i. \tag{4.16}$$

Evidently the pair c_i, ϕ_i for each i are action-angle variables. Consequently, we have

$$\dot{\phi}_i = \left(\frac{\partial h^*}{\partial c_i}\right) = \omega_i(c_1,c_2). \tag{4.17}$$

Since the c_i are integrals of motion (h^* is independent of the ϕ_i thanks to the normal form algorithm), the frequencies ω_i are independent of the time. Consequently, Eq. (4.17) can be integrated directly to give

$$\phi_i = \omega_i t + \phi_i^0. \tag{4.18}$$

Thus, combining (4.16) through (4.18), we find that the motion of the q_i, p_i is periodic with the frequencies ω_1 and ω_2. Since h^{old} and h^* are related by a canonical transformation, it follows that the motion described by h^{old} must be quasi-periodic with the two fundamental frequencies ω_1 and ω_2. In particular, orbits for which the ratio ω_2/ω_1 is a rational number will be closed, and therefore will be completely periodic.

Suppose we consider all the orbits for which the energy has a particular *set value* and for which the fundamental frequency ratio is rational, $\omega_2/\omega_1 = m/n$. Since h^* and the ω_i depend only on c_1 and c_2, these two conditions determine the values of the integrals c_1 and c_2. However, ϕ_1^0 and ϕ_2^0 are undetermined. Consequently, we expect that those periodic orbits having a particular energy and frequency ratio will form a two-dimensional surface in phase space. However, direct numerical integration of orbits for the Størmer problem shows that this is not the case. Instead, one finds that for a fixed energy, there are only a finite number of orbits having a specified rational value for ω_1/ω_2. Put another way, for a fixed energy, the closed (and therefore periodic) orbits in phase space are isolated curves, and do not form a two-dimensional surface.[16] It follows that the normal form process that we have described in this paper must be divergent for the Størmer problem.

We are currently working on a different normal form procedure in order to overcome this difficulty.[14] However, we wish to point out here that the apparently formal expres-

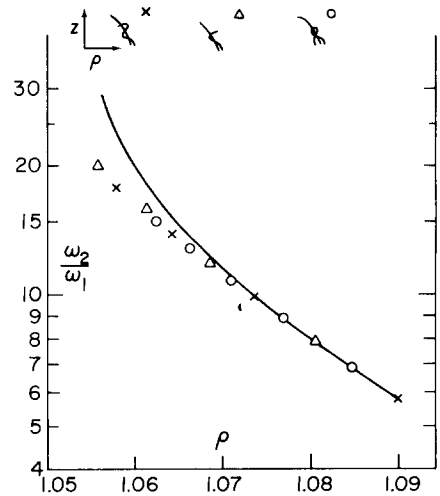

FIG. 5. A continuation of Fig. 4 showing analogous results for those orbits whose initial conditions are such that $\rho > 1$.

TABLE I. A comparison of initial conditions for periodic orbits estimated using the normal form (4.21) in (4.17) and initial conditions obtained exactly by the numerical integration of trajectories. There are no orbits with rotation number less than 10 for this energy ($h = 0.01$).

Rotation number ω_2/ω_1	Estimated q_2	Numerically determined q_2
10	0.1413772	0.1413772
11	0.1349294	0.1349264
12	0.1294129	0.1294038
13	0.124624	0.124608
14	0.120415	0.120393
15	0.116678	0.116650
16	0.113330	0.113298
17	0.110308	0.110272
18	0.107561	0.107523
19	0.105050	0.105009
20	0.102743	0.102700
22	0.098638	0.098592
24	0.095083	0.095035
27	0.090542	0.090493
30	0.086722	0.086672
33	0.083447	0.083397
36	0.080596	0.080547
39	0.078084	0.078036

sions (4.15) and (4.17) for the ω_i still are useful. For when periodic orbits are located numerically and ω_2/ω_1 is then computed from (4.15) and (4.17) using the numerically determined initial conditions, one finds that the ratio is in fact nearly rational. Figures 4 and 5 illustrate how well this works. From these figures we see that the results are quite accurate for nearly equatorial orbits, but are worse for orbits whose mirror points are further down the guiding field line. Moreover, we remark that for initial values of ρ satisfying $0.9383 < \rho < 1.082$, which includes most of the interesting region of the figures, the value of q_1 exceeds $3\sqrt{3}/16$ somewhere on the trajectory, and consequently we expect trouble on that basis alone. Thus, the accuracy of our results is much better than we have any reason to expect.

We devote the remainder of this section to a discussion of orbits in the simple model mirror machine whose magnetic field is given by (4.1). In particular, we shall focus our attention on those orbits for which $p_\phi = 0$. These orbits do not encircle the central axis of the mirror machine, but instead pass continually through it. Hence, one cannot use energy conservation arguments to preclude their extending arbitrarily far down the length of the machine. After suitable scaling, we arrive at the simple Hamiltonian

$$h^{\text{old}} = \tfrac{1}{2}p_1^2 + \tfrac{1}{2}(p_2^2 + q_2^2) + \tfrac{1}{2}q_1^2q_2^2 + \tfrac{1}{8}q_1^4q_2^2. \tag{4.19}$$

Here q_1 is proportional to z and q_2 is proportional to ρ.

In this case the Hamiltonian is considerably simpler than that for the Størmer problem. Consequently, we were able to carry out the normal form algorithm to higher order within the computer time available. The result through terms of ninth order is

$$h^{\text{new}} = \tfrac{1}{2}p_1^2 + c_2 + \tfrac{1}{2}c_2q_1^2 + \tfrac{1}{16}c_2^2q_1^2 + (\tfrac{1}{16}c_2^3q_1^2 - \tfrac{5}{48}c_2^2q_1^4). \tag{4.20}$$

As before, we may now perform a variant of the normal form algorithm a second time to arrive at a final Hamiltonian which only depends on the variables c_1 and c_2. The result is

$$h^*(c_1,c_2) = c_2 + (c_2 + \tfrac{1}{8}c_2^2 + \tfrac{1}{8}c_2^3)^{1/2}c_1 - \tfrac{5}{32}(1 + \tfrac{1}{8}c_2 + \tfrac{1}{8}c_2^2)^{-1}c_2c_1^2 + \cdots. \tag{4.21}$$

A numerical study of the periodic orbits for a fixed value of the energy shows that they are again isolated, and do not form a two dimensional surface. In particular, various periodic orbits have initial conditions of the following form: $q_1 = 0; p_2 = 0; q_2$ variable and taking on various discrete values; and p_1 determined by energy conservation once q_1, q_2, and p_2 have been specified. Table I shows the initial value of q_2 for various periodic orbits in the case for which the equation $h^{\text{old}} = 0.01$ fixes the energy. Each integer in the column labeled "rotation number" is defined to be the number of oscillations undergone by the variable q_2 during one oscillation of q_1. It corresponds physically to the number of gyrations a charged particle makes during one complete mirroring cycle, and should ideally equal the ratio ω_2/ω_1 computed by the normal form algorithm. This is very nearly the case. The column labeled "estimated q_2" is the value of q_2 computed by requiring that the ratio ω_2/ω_1 computed using (4.21) and (4.17) be exactly equal to the rotation number in question. (We have, of course, also required that $h^{\text{old}} = 0.01$, $p_2 = 0$, and $q_1 = 0$.)

The agreement between the columns "estimated q_2" and "numerically determined q_2" is remarkably good, and illustrates the utility of the normal form approach even though the series employed must ultimately be divergent. We observe that the agreement for the model machine problem is much better than it was for the Størmer problem. This is because in (4.20) the ratio of the coefficient of the terms involving q_1^2 is smaller by a factor of c_2 than the corresponding ratio for (4.6). This circumstance can be traced to the fact that (4.19) contains no terms of order 3 [whereas (4.4) does], and is a special feature of orbits with $p_\phi = 0$.

Because of its simplicity, the model mirror machine problem is an ideal context in which to examine the integral of motion produced by the normal form algorithm. We have computed i^{old} through terms of ninth order using (1.15) with $i^{\text{new}} = c_2$. The entire expression is too lengthy to record, and is best transferred directly from one computer program to another. The first few terms are given by

$$\begin{aligned} i_2^{\text{old}} &= \tfrac{1}{2}(p_2^2 + q_2^2), \\ i_3^{\text{old}} &= 0, \\ i_4^{\text{old}} &= \tfrac{1}{8}(p_1^2p_2^2 - p_1^2q_2^2 - 2p_2^2q_1^2 + 2q_1^2q_2^2 + 4p_1p_2q_1q_2). \end{aligned} \tag{4.22}$$

Examination of these terms and those of higher order shows that i^{old} contains within it all the terms in the power series expression for the magnetic moment, $E_\perp/B$:

$$E_\perp/B = \tfrac{1}{2}p_2^2/(1 + \tfrac{1}{2}q_1^2) + \tfrac{1}{2}q_2^2 + \tfrac{1}{4}q_1^2q_2^2 = \tfrac{1}{2}(p_2^2 + q_2^2) + \tfrac{1}{4}(q_1^2q_2^2 - p_2^2q_1^2) + \cdots. \tag{.4.23}$$

There are also additional terms in i^{old} beyond these. They represent corrections required to make up the *complete adiabatic magnetic moment expansion.*

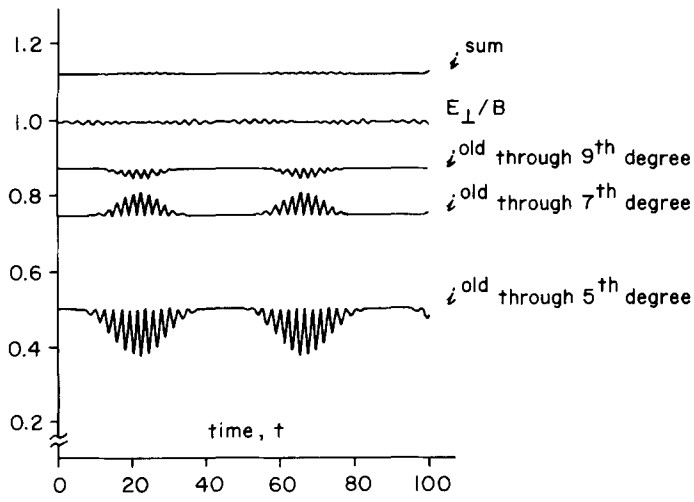

FIG. 6. The quantities i^{old} through various orders, $E_\perp/B$, and i^{sum} as functions of time for the orbit with initial conditions $q_1 = q_2 = 0$ and $\dot{q}_1 = 0.0707$, $\dot{q}_2 = \frac{1}{10}$. The mirror point is at $q_1 \simeq 1.0$, and the magnetic field strength at the mirror point is 1.5 times its value at $q_1 = 0$. The vertical scale is arbitrary, and to separate the curves a different constant has been added to each.

Since i^{old}, as a series, contains all the terms in (4.23), it is reasonable to sum these terms explicitly. When this is done, we will obtain a quantity, denoted by i^{sum}, which contains $E_\perp/B$ exactly plus all terms of the remainder of the complete adiabatic magnetic moment expansion through the highest degree to which i^{old} is computed. The quantity i^{sum} may be expected to be more nearly constant than i^{old}. More explicitly, through terms of order nine we write

$$i^{sum} = i^{old} - \tfrac{1}{2}p_2^2(1 - \tfrac{1}{2}q_1^2 + \tfrac{1}{4}q_1^4 - \tfrac{1}{8}q_1^6) + \tfrac{1}{2}p_2^2/(1 + \tfrac{1}{2}q_1^2). \tag{4.24}$$

Note that the series for the magnetic moment $E_\perp/B$ given in (4.23) diverges for $|q_1| > (2)^{1/2}$. We therefore do not expect that the function i^{old} truncated at a finite degree will be a good integral of motion for orbits which mirror beyond $|q_1| = (2)^{1/2}$. But this does not rule out the possibility of i^{sum} being very nearly constant, when truncated at high degree.

To examine the constancy of i^{old} and i^{sum} as integrals of motion, we have integrated numerically the equations of motion. Figure 6 shows graphs of i^{old}, taken through 5th, 7th, and 9th order respectively, as a function of time over an orbit. The orbit starts at $q_1 = 0$ at $t = 0$, mirrors at $t \simeq 22$, recrosses the median plane $q_1 = 0$ at $t \simeq 44$, mirrors again at $t \simeq 66$, etc. It is evident that the constancy of i^{old} improves as more terms in the series are included. (We will see later, however, that the series must ultimately diverge, so that improvement cannot continue indefinitely.) Also shown is the quantity i^{sum}, computed from (4.24) with i^{old} taken through ninth degree. It remains remarkably constant over the entire orbit. Finally, the quantity $E_\perp/B$ is plotted to demonstrate that i^{old} is superior to the magnetic moment except at mirror points, and that i^{sum} is superior everywhere. Results of several numerical integration runs for a range of initial conditions show that this behavior holds in general.

We close this section with the presentation of numerical evidence that the model mirror machine problem, like the Størmer problem, is insoluble. What it means for a classical mechanics problem to be "insoluble" and how this can come about has been described in detail elsewhere.[4] In essence, it means that there are, in fact, no analytic functions of the p's and q's which are integrals of motion and therefore satisfy (1.17). Consequently, trajectories in phase space are not confined to lie on analytic hypersurfaces. Instead, they may wander in a very complicated way, and are sufficiently complex so as to preclude their explicit representation. It also follows that the complete adiabatic magnetic moment expansion is divergent. Finally, it can also be shown that any of the standard methods of classical mechanics, such as perturbation series or solution of the Hamiltonian-Jacobi equation, must also fail. In particular, there is no known way of predicting the long-term behavior of trajectories; and in the case of mirror machines, long-term containment cannot be mathematically guaranteed.

One method of testing numerically for the existence of integrals of motion is to plot the values of some independent pair of variables each time a trajectory in phase space crosses the median plane $q_1 = 0$. If the result of such a plot is a collection of points which have the appearance of lying on a smooth curve, then the existence of an integral is suggested, although not proved. By contrast, if no such regularity occurs, the existence of an integrtal is ruled out. Figure 7 shows variations in i^{old} (through ninth order) plotted against p_2 for successive median plane crossings. It is evident that these points are scattered, and the existence of an integral of motion is precluded. Thus, the model mirror machine, despite its simplicity, is insoluble. More detail may be found in Ref. 3.

5. CONCLUDING SUMMARY

In Secs. 1 and 2 a partial exploration was made of the

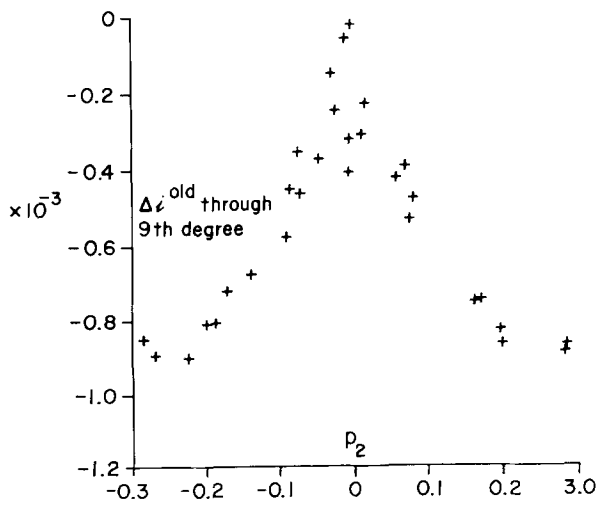

FIG. 7. Changes in the value of i^{old} through ninth degree plotted versus p_2 for successive median plane crossings. The initial conditions for this orbit are $q_2 = 0.299$, $p_2 = 0$, $q_1 = 0$, and p_1 determined by the energy condition $h^{old} = \frac{1}{10}$.

effect of canonical transformations on mirror machine Hamiltonians, and a certain normal form was shown to be always possible. In Sec. 3 it was shown that this normal form led to the existence of a formal integral. Section 4 treated two examples of mirror machines, and normal form methods were used to obtain expressions for bounce frequencies and the location of periodic orbits. The lengthy algebraic calculations required were performed by computer. Good agreement was found between analytical and numerical results for periodic orbits. In addition, it was shown that the integral of motion produced by the normal form method is in fact a series for the complete adiabatic magnetic moment expansion. Finally, it was shown that, like the Størmer problem, the simple model mirror machine does not possess an analytic third integral of motion. Therefore its complete adiabatic magnetic moment expansion is divergent, phase space trajectories are not confined to lie on analytic hypersurfaces, and the problem, despite its apparent simplicity, is insoluble. In particular, there is no mathematical guarantee of long-term containment.

ACKNOWLEDGMENT

The first author is indebted to the Institut des Hautes Etudes Scientifiqes where this work was begun. He wishes to thank Professor N. Kuiper and Professor L. Michel for their fine hospitality. We are both indebted to the University of Maryland Computer Science Center, which provided the computer time essential to this research, and to the National Science Foundation for their partial support.

APPENDIX

Suppose that the frequencies α_i in (2.2) exhibit some degree of incommensurability.[17] In particular, we assume that there are l and only l linearly independent equations of commensurability between the frequencies $\alpha_{m+1},...,\alpha_n$. We write these relations in the form

$$\sum_{j=m+1}^{n} \mathcal{M}_{ij}\alpha_j = 0, \quad i = 1,...,l. \tag{A1}$$

Equivalently, we write

$$\mathcal{M}\alpha = 0 \tag{A2}$$

where α is a vector with entries $\alpha_{m+1}\cdots\alpha_n$, and $\mathcal{M}$ is an $l \times (n-m)$ matrix with integer coefficients, rank l, and indexed so that $m+1 \leqslant j \leqslant n$.

Using a notation similar to (3.2a), we write

$$c = \sum_{j=m+1}^{n} \alpha_j c_j, \tag{A3}$$

where

$$c_j = \tfrac{1}{2}(p_j^2 + q_j^2). \tag{A4}$$

We also write

$$C = \sum_{j=m+1}^{n} \alpha_j C_j. \tag{A5}$$

It is evident that the various Lie operators C_j mutually commute, since they involve differentiation with respect to different variables. Also, from (2.17) they are all antihermitian. Finally, they each map $\mathcal{P}_k$ on to itself for each value of k. It follows that there exists a basis in which the C_j are all simultaneously diagonal.

We next assert that the eigenvalues of each C_j are the integers multiplied by i. To see this, we introduce monomials in the single pair q_s, p_s defined by

$$|j_s;m_s) = (p_s + iq_s)^{j_s+m_s}(p_s - iq_s)^{j_s-m_s}. \tag{A6}$$

Here, as before, the quantities j_s and m_s are integral or half integral. We find, after simple computation, that

$$C_s|j_s;m_s) = -2im_s|j_s;m_s). \tag{A7}$$

Since any monomial in the various variables can be built of products of the form (A6), and since these monomials obviously form a basis, our assertion is proved.

Let $|\lambda_{m+1}\cdots\lambda_n\rangle$ denote an eigenvector of the various C_j. It is constructed from products of monomials of the form (A6). We have

$$C_j|\lambda_{m+1}\cdots\lambda_n\rangle = i\lambda_j|\lambda_{m+1}\cdots\lambda_n\rangle, \tag{A8}$$

where the λ's are integers. It follows from (A5) and (A8) that

$$C|\lambda_{m+1}\cdots\lambda_n\rangle = \left(i\sum_{j=m+1}^{n}\alpha_j\lambda_j\right)|\lambda_{m+1}\cdots\lambda_n\rangle. \tag{A9}$$

Since C is antihermitian, we may rewrite (A9) in the form

$$\langle\lambda_{m+1}\cdots\lambda_n|\, C = \left(i\sum_{j=m+1}^{n}\alpha_j\lambda_j\right)\langle\lambda_{m+1}\cdots\lambda_n|\,. \tag{A10}$$

Now take the scalar product of both sides of (A10) with the vector $|\,h^{\text{new}}\rangle$. We find, using (3.19), the result

$$\begin{aligned} 0 &= \langle\lambda_{m+1}\cdots\lambda_n|\,C\,|h^{\text{new}}\rangle \\ &= \left(i\sum_{j=m+1}^{n}\alpha_j\lambda_j\right)\langle\lambda_{m+1}\cdots\lambda_n|\,h^{\text{new}}\rangle. \end{aligned} \tag{A11}$$

Thus, for $|\lambda_{m+1}\cdots\lambda_n\rangle$ to appear in h^{new}, the eigenvalues λ_j must necessarily satisfy the relation

$$\sum_{j=m+1}^{n}\alpha_j\lambda_j = 0. \tag{A12}$$

By the hypothesis (A1) or (A2), there are only l linearly independent relations of the form (A12) with integer coefficients. It follows that for each set of λ's obeying (A12) there must be other coefficients $\beta_1^\lambda\cdots\beta_l^\lambda$ such that

$$\lambda_j = \sum_{i=1}^{l}\beta_i^\lambda\mathcal{M}_{ij} \tag{A13}$$

or, using matrix and vector notation,

$$\lambda = \tilde{\mathcal{M}}\beta^\lambda. \tag{A14}$$

We are almost done. Let $\gamma_{m+1}\cdots\gamma_n$ be a set of real numbers such that the vector γ with entries γ_j is an eigenvector of $\mathcal{M}$ with eigenvalue zero,

$$\mathcal{M}\gamma = 0. \tag{A15}$$

By the nature of $\mathcal{M}$, there will be $n-m-l$ such vectors γ which are linearly independent. Using (A14) and (A15), we find that the scalar product between any λ obeying (A12) and any γ obeying (A15) vanishes,

$$(\lambda,\gamma) = (\tilde{\mathcal{M}}\beta^\lambda,\gamma) = (\beta^\lambda,\mathcal{M}\gamma) = 0. \tag{A16}$$

Now define operators C_γ by the rule

$$C_\gamma = \sum_{j=m+1}^{n} \gamma_j C_j. \tag{A17}$$

There will $n-m-l$ such operators which are linearly independent. They evidently have the property that

$$C_\gamma |\lambda_{m+1}\cdots\lambda_n\rangle = i(\lambda,\gamma)|\lambda_{m+1}\cdots\lambda_n\rangle = 0 \tag{A18}$$

when (A12) holds. Since the only vectors $|\lambda_{m+1}\cdots\lambda_n\rangle$ occurring in the expansion of h^{new} are those for which (A12) holds, we must also have

$$C_\gamma h^{\text{new}} = 0. \tag{A19}$$

It follows that the $n-m-l$ functions i_γ^{new} defined by $i_\gamma^{\text{new}} = c_\gamma$, where

$$c_\gamma = \sum_{j=m+1}^{n} \gamma_j c_j \tag{A20}$$

are integrals of motion for h^{new}.

We note that, in this notation, c_α is the integral already found in Sec. 3. We also remark that in the extreme case in which $l=0$, i.e., all the α's are irrational in a way as to be completely incommensurate, then each c_j is an integral of motion for h^{new}.

[1]R. De Vogelaere, *Contributions to the Theory of Nonlinear Oscillations*, edited by S. Lefshetz (Princeton U.P., Princeton, New Jersey, 1958), Vol. 4, p. 53.

[2]A.J. Dragt, "Trapped orbits in a magnetic dipole field," Rev. Geophys. Space Phys. **3**, 225 (1965); correction, Geophys. Space Phys. **4**, 112 (1966).

[3]J.M. Finn, "Integrals of canonical transformations and normal forms for mirror machine Hamiltonians," Ph.D. thesis, Univ. Maryland (1974).

[4]A.J. Dragt and J. M. Finn, "Insolubility of trapped particle motion in a magnetic dipole field," J. Geophys. Res. **81**, 2327 (1976).

[5]Island structure and separatrices in the context of wave–particle interactions are described in, for example, G.R. Smith, "Overlap of bounce resonances and the motion of ions in a trapped-ion mode," Phys. Rev. Lett. **38**, 970 (1977); C.F. Karney and A. Bers, "Stochastic ion heating by a perpendicularly propagating electrostatic wave," Phys. Rev. Lett. **39**, 550 (1977).

[6]R. Cohen, G. Rowlands, and J. Foote, "Nonadiabaticity in mirror machines," Phys. Fluids **21**, 627 (1978); H.E. Taylor, and R.J. Hastie, "Nonadiabatic Behavior of Radiation-Belt Particles," Cosmic Electrodynamics **2**, 211 (1971); J.E. Howard, "Nonadiabatic Particle Motion in Cusped Magnetic Fields," Phys. Fluids **14**, 2378 (1971).

[7]See, for examle J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton U. P., Princeton, NJ, 1973); B.V. Chirikov, "A universal instability of many dimensional oscillator systems," Phys. Rep. (in press).

[8]A. Deprit, J. Henrard, J. Price, and A. Rom, "Birkhoff's Normalization," Celestial Mech. **1**, 222 (1969); M.V. Berry and M. Tabor, "Calculating the bound spectrum by path summation in action-angle variables," J. Phys. A: Math. Gen. **10**, 371 (1977); I.C. Percival and N. Pomphrey, "Vibrational quantization of polyatomic molecules, "Molec. Phys. **31**, 97 (1976); see also Ref. 17.

[9]A.J. Dragt and J.M. Finn, "Lie series and invariant functions for analytic symplectic maps," J. Math. Phys. **17**, 2215 (1976).

[10]See, for example, P.R. Halmos, *Finite Dimensional Vector Spaces* (Van Nostrand, Princeton, NJ, 1958), p. 127.

[11]For those familiar with Lie Algebras, we remark that the associated Lie operators constitute the *adjoint* representation of the underlying Poisson bracket Lie Algebra; hence the notation. See, for example, M. Hausner and J.T. Schwartz, *Lie Groups; Lie Algebras* (Gordon & Breach, New York, 1968). Computation by the reader may be simplified by noting that $\text{ad}(q_i^2) = 2q_i(\partial/\partial p_i)$, etc.

[12]See, for example, E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961), p. 361.

[13]Note that although (2.19) always has a solution, the solution is not unique. Given a solution, one can always produce a new solution simply by adding on a function in the null space of H_2. The implications of this freedom of choice have not been fully explored. In our calculations to date we have imposed the further requirement that each f_k be orthogonal to the null space of H_2. When considering the question of uniqueness, it should also be noted that the choice of scalar product, which in turn defines H_2^+ by (2.12), is also not unique. This degree of freedom also has not been fully explored.

[14]For a study of those orbits which eventually close on themselves, and hence are periodic, a different procedure may be more useful. It is known that such orbits correspond to fixed points of a Poincare surface of section map or its iterates, and that these fixed points are in turn the locale or origin of island or homo–heteroclinic behavior. (See Ref. 4.) A preliminary step toward the construction of integrals in these so called "resonance" cases has been made in Ref. 15.

[15]G. Contopoulos and L. Vlahos, "Integrals of motion and resonances in a pure dipole magnetic field," J. Math. Phys. **16**, 1469 (1975). For related work, see "Resonance Cases and Small Divisors in a Third Integral of Motion I–III," Astron. J. **68**, 763 (1963); **70**, 817 (1965); **71**, 687 (1966); "Tables of the Third Integral," Astrophys. J. Suppl. **13**, 122 (1966).

[16]See E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, 4th ed. (Cambridge U.P., Cambridge, England, 1960), p. 396.

[17]F. Gustavson, "On Constructing Formal Integrals of a Hamiltonian System near an equilibrium point," Astron. J. **71**, 670 (1966).

Zeros of the grand partition function: One-dimensional Widon–Rowlinson lattice gas[a)]

L. K. Runnels and John Runnels
Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana 70803

(Received 19 June 1979; accepted for publication 21 August 1979)

The distribution of zeros of the partition function is studied for a one-dimensional symmetric two-component lattice gas. It is shown explicitly that the (complex) activities of the two components need not in general be of equal magnitude for the partition function to vanish, although such is the case if the geometric mean activity is sufficiently high. The effect of boundary conditions on the distribution of zeros is also studied.

1. INTRODUCTION

Lattice versions of the two-component model introduced by Widom and Rowlinson have proved their theoretical value in enlarging our understanding of the statistical thermodynamics of interacting systems and phase transitions.[1] The very first role played by these models[2] was the demonstration of the versatility of the Peierls contour technique of proving the existence of a phase transition—originally developed for Ising spin systems. In the basic two-component lattice model of Ref. 2, each component excludes the other from its "sphere" of nearest-neighbor sites, while not interacting at all with other molecules of the same type (except that multiple occupancy of any site is always forbidden). It was shown that in two or more dimensions, if the activities of the two components are (real and) equal, as that common activity is increased a value is reached at which a demixing phase transition occurs. Alternatively, we can say that above the critical activity the equilibrium state is not unique but is dependent on the boundary conditions.

The requirement that the two activities be equal is crucial to the success of the Peierls argument, just as the assumption of zero magnetic field is crucial to its success when applied to Ising spin systems. Just because the requirement is needed to make the proof work does not, of course, prove the converse of the theorem. In the Ising case, however, we also have the Yang–Lee "circle theorem"[3] which insures analytic thermodynamics unless the (complex) exponential of the magnetic field is of unit magnitude—consequently, if the field is real, it must be zero for a phase transition to occur.

We have a fairly analogous result[4] for the lattice Widom–Rowlinson gas, in which the two independent variables are taken to be the ratio of the two activities and their (geometric) mean. The ratio than plays a role analogous to that of the magnetic field in the Ising case and the mean activity is the analog of the temperature. While the Yang–Lee circle theorem holds at all temperatures, we were only able to prove the corresponding result (nonanalyticity implies modulus of ratio equals one) for the lattice Widom–Rowlinson case for sufficiently high mean activity (sufficiently low temperature in magnetic language).

It was not clear whether the failure of the proof at low activities was a shortcoming in the technique, or whether it has greater significance. What is expected physically is that, for *real* values of the activities of the highly symmetric model, any nonanalyticity should occur at ratio one. However, permitting the ratio to assume complex values could conceivably result in a locus of singularities that crosses the real axis only at the point one—but that locus might not be a circle.

Consequently, this study of the one-dimensional version of the model was undertaken in hopes of shedding some light on this question with an essentially exactly soluble model. Additionally, it was found that this model also gives some information on the effect of boundary conditions on the distribution of zeros of the partition function.

2. THE MODEL

The one-dimensional, two-component Widom–Rowlinson model on a lattice is described by the transfer matrix

$$\mathbf{T} = \begin{pmatrix} 1 & x^{1/2} & y^{1/2} \\ x^{1/2} & x & 0 \\ y^{1/2} & 0 & y \end{pmatrix},$$

where x and y are the activities, respectively, of particles of type 1 and 2. The transfer matrix reflects the exclusion of adjacent unlike particles and the absence of any interaction between neighboring like particles or between a void site and any type of site. If $\langle\psi_p|$ and $\langle\psi_f|$ stand for the vectors

$$\langle\psi_p| = (1,1,1)$$

and

$$\langle\psi_f| = (1,\sqrt{x},\sqrt{y}),$$

then the grand partition functions for L sites with periodic ("p") and tree ("f") boundary conditions become

$$\Xi_p(x,y) = \langle\psi_p|\mathbf{T}^L|\psi_p\rangle,$$

$$\Xi_f(x,y) = \langle\psi_f|\mathbf{T}^{L-1}|\psi_f\rangle.$$

In either case, for real positive activities x and y, evaluating the partition function becomes a matter of determining the dominant eigenvalue of $\mathbf{T}$, at least in the thermodynamic limit $L\to\infty$. In order to study zeros of the partition function however—which are not in general restricted to real and

[a)]Supported in part by the National Science Foundation Grant No. CHE76-11253.

 0022-2488/79/122661-04$01.00

positive activities—it is necessary to include *all* eigenvalues and match boundary conditions. We thereby obtain expressions of the form

$$\Xi_\alpha = a_\alpha \lambda_1^L + b_\alpha \lambda_2^L + c_\alpha \lambda_3^L, \tag{1}$$

where the λ_i's are the eigenvalues of T—functions of x and y—and where the label α can be either p or f. The straightforward approach is then to study the algebraic equation

$$\Xi_\alpha(x,y) = 0,$$

which can be solved in favorable cases.

A more powerful method for periodic boundary conditions is based on the observation[5] than $\ln\Xi_p$ can fail to be analytic in its dependence on x and y only when the dominant eigenvalue fails to be unique (again, in the thermodynamic limit). This leads to equations of the sort

$$|\lambda_1| = |\lambda_2| > |\lambda_3|$$

and other permutations of the subscripts.

The secular equation for T, with roots λ_1, λ_2, and λ_3, is

$$\lambda^3 - (1 + x + y)\lambda^2 + xy\lambda + xy = 0 \tag{2}$$

and shows that the two activities actually enter most naturally through their arithmetic and geometric means. We will at times use the definitions

$$xy = z^2$$

and

$$1 + x + y = \xi$$

so that Eq. (2) becomes

$$\lambda^3 - \xi\lambda^2 + z^2\lambda + z^2 = 0. \tag{3}$$

We study two special cases to explore (I) boundary condition effects on the zeros and (II) the effect of the ratio x/y on the zeros. To study (I) we impose the simplifying condition that the ratio x/y be one and to study (II) we impose the simplification of periodic boundary conditions and the requirement that the product $xy = z^2$ be real.

3. CASE (I): EQUAL ACTIVITIES

For equal activities it is readily learned that the three roots of Eq. (2) are

$$\lambda_1 = x, \tag{4a}$$

$$\lambda_2 = \tfrac{1}{2}[(1 + x) + (x^2 + 6x + 1)^{1/2}], \tag{4b}$$

$$\lambda_3 = \tfrac{1}{2}[(1 + x) - (x^2 + 6x + 1)^{1/2}], \tag{4c}$$

and that the expansion coefficient a_f in Eq. (1) vanishes for *free* boundary conditions. The vanishing of Ξ_f then leads to

$$\lambda_3/\lambda_2 = (-1)^{1/(L-1)} \tag{5}$$

in the thermodynamic limit $L \to \infty$. Interpreting this last equation as an equation for the (common) activity x, it is easy to show that x must be real and negative and in the range

$$-3 - 2\sqrt{2} \leqslant x \leqslant -3 + 2\sqrt{2}. \tag{6}$$

This follows from noticing that Eq. (5) means that $|\lambda_2/\lambda_3| = 1$, which in turn requires $(x^2 + 6x + 1)^{1/2}$ to be pure imaginary.

Now, for periodic boundary conditions, all three coefficients in Eq. (1) are nonzero, and λ_1 must also be included in the analysis. The easiest way to do this is by way of the Katsura approach.[5] One of the three equations is

$$|\lambda_2| = |\lambda_3| > |\lambda_1|. \tag{7}$$

It may be shown that Eq. (7) is satisfied as long as x is real and in the range

$$-1 \leqslant x \leqslant -3 + 2\sqrt{2}, \tag{8}$$

which is *part* of the line segment represented by Eq. (6).

The remaining equations

$$|\lambda_1| = |\lambda_2| \geqslant |\lambda_3|$$

and

$$|\lambda_1| = |\lambda_3| \geqslant |\lambda_2|$$

allow x to be complex. Its possible values may be shown to lie on the curve

$$x = \frac{e^{i\phi} - 1}{e^{-i\phi} + 1} \tag{9}$$

parameterized by $\phi \in [\pi/2, 3\pi/2) - \{\pi\}$ (see Fig. 1).

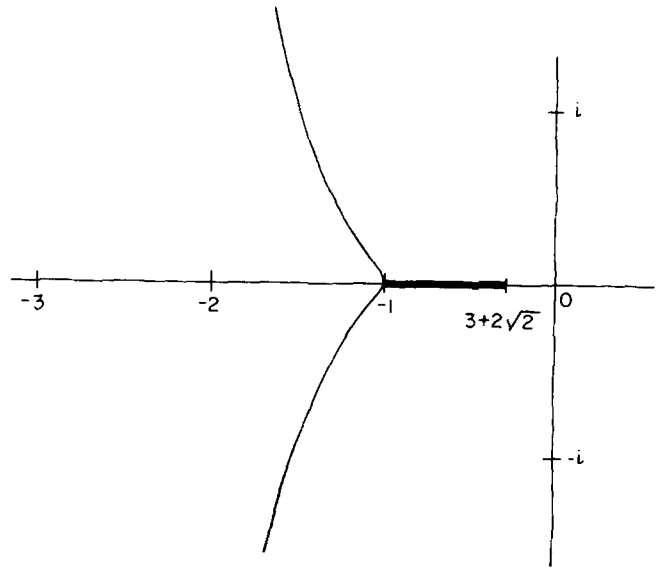

FIG. 1. Locus of zeros, equal activities, and periodic boundary conditions. With free boundary conditions the complex zeros disappear, while negative zeros extend further (to $-3 - 2\sqrt{2}$).

In a certain sense we can say that the periodic boundary conditions render the system more interacting and more capable of cooperative behavior. Hence, compared to the zeros for free boundaries, some of the zeros have moved off the negative real axis. There is, of course, no phase transition in this one-dimensional system—that would require that some of the zeros lie arbitrarily close to the positive real axis.

4. CASE (II): UNEQUAL ACTIVITIES

Now we address the following question: Is it possible for $\Xi_p(x,y)$ to vanish if $x \neq y$? Intuition certainly expects a negative answer if x and y are both restricted to real, positive (i.e., physical) values. This intuition is based on the symmetry operation of exchanging type 1 particles for type 2 everywhere, analogous to spin flips in an Ising model in zero field.

The original Yang–Lee work[3] showed that the Ising partition function with complex argument $e^{-\beta H}$ cannot vanish unless $|e^{-\beta H}| = 1$—so if H is real it must be zero.

Inasmuch as there is an analogous circle theorem for the Widom–Rowlinson model[4] (any number of dimensions) for sufficiently high activities, it is natural to wonder if there is a corresponding theorem requiring $|x/y|$ to be one *any time* that $\Xi_p(x,y)$ vanishes.

Actually, that statement is false, as the present one-dimensional example shows. Besides periodic boundaries, the other simplification we make is to study only pairs of complex variables x and y such that $xy = z^2$ with z real. This reduces by one the number of variables to manipulate since $\mathrm{Arg}(x) = -\mathrm{Arg}(y)$. It develops, however, that this still provides sufficient latitude to find pairs (x,y) of this sort where $x \neq y^*$ but $\Xi_p(x,y) = 0$. (Clearly, if $x = y^*$, then $|x/y| = 1$.) The proof consists of constructing the pairs of points with the just named property $x \neq y^*$ but $\Xi_p(x,y) = 0$. We also learn conditions (on z) that permit such asymmetric zeros.

Notice that the secular Eq. (3) implies the three symmetric equations

$$\lambda_1\lambda_2\lambda_3 = -z^2, \tag{10a}$$

$$\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = z^2, \tag{10b}$$

$$\lambda_1 + \lambda_2 + \lambda_3 = \xi. \tag{10c}$$

Our goal is to produce three complex numbers λ_i satisfying these three equations with the additional restraint that two of them (say λ_1 and λ_2) have the same magnitude

$$|\lambda_1| = |\lambda_2| = r > 0. \tag{11}$$

We must also impose the condition that ξ not be real for $\mathrm{Im}(\xi) = 0$ implies that $\mathrm{Im}(x + y) = 0$, which means that $x = y^*$ since $\mathrm{Arg}(x) = -\mathrm{Arg}(y)$.

We thus look for solutions of Eq. (10) of the form

$$\lambda_1 = r\,e^{i(\psi - \chi)}, \tag{12a}$$

$$\lambda_2 = r\,e^{i(\psi + \chi)}, \tag{12b}$$

with ψ and χ real and $\psi \neq 0$, assuming z real and ξ not real. Clearly, from Eq. (10a) we have

$$\lambda_3 = -(z/r)^2 e^{-2i\psi} \tag{12c}$$

and then from Eq. (10b) it follows that

$$\lambda_1\lambda_2 = z^2 - \lambda_3(\lambda_1 + \lambda_2) = z^2(1 + r^{-1}e^{-i(\psi+\chi)} + r^{-1}e^{-i(\psi-\chi)}). \tag{13a}$$

Since $\lambda_1\lambda_2 = r^2 e^{2i\psi}$, the equation

$$r^3 e^{3i\psi} = z^2(r\,e^{i\psi} + 2\cos\chi) \tag{13b}$$

replaces Eq. (10b).

Now the imaginary part of Eq. (13b) is independent of χ:

$$r^2\sin 3\psi = z^2\sin\psi.$$

Since we are interested in nonzero values of ψ, we may divide both sides of this last equation by $\sin\psi$, using $\sin 3\psi = 3\sin\psi - 4\sin^3\psi$, to obtain

$$4\sin^2\psi = 3 - (z/r)^2. \tag{14}$$

Since ψ is to be real, there is implied the condition

$$z/r < 3^{1/2}. \tag{15}$$

Next the real part of Eq. (13b), i.e.,

$$r^3\cos 3\psi = z^2 r\cos\psi + 2z^2\cos\chi = r^3\cos\psi(4\cos^2\psi - 3),$$

may be solved for $\cos\chi$ using Eq. (14). The result is simply

$$z^2\cos\chi = -r^3\cos\psi. \tag{16}$$

Requiring $0 < \cos^2\chi < 1$ yields another conditions on z and r:

$$r^4(r^2 + z^2) < 4z^4. \tag{17}$$

Equations (15) and (16) give, for a specified z, restrictions on the possible values of r. A final restriction is implicit in Eq. (12c) since $|\lambda_3|$ must be less than $|\lambda_1| = |\lambda_2| = r$:

$$z^2/r^2 < r$$

or

$$r^3 > z^2. \tag{18}$$

To summarize, if we use the common magnitude of the two largest eigenvalues to parameterize our description, that parameter r must simultaneously satisfy the lower bound

$$r > \max(3^{-1/2}z, z^{2/3}) \tag{19}$$

and the upper bound implicit in Eq. (17). As z grows larger the upper bound is of order $z^{1/2}$ while the lower bound is of order $z^{2/3}$. It follows that there is a number z_0 above which all conditions cannot be met, and therefore above which the circle theorem result must hold. It is not difficult, in fact, to discover that $z_0 = 3^{3/2} = 5.196...$

The program, then, for $z < z_0$ is to choose first any r in the range defined by the lower boundary [Eq. (19)] and the upper bound [Eq. (17)]. Using this chosen value for r, Eq. (14) defines a real value of ψ and then Eq. (16) defines χ—also real. We then work backwards: Equations (12) give the eigenvalues λ_i corresponding to activities x and y, which are given by $xy = z^2$ and $1 + x + y = \Sigma\lambda_i$. Since the two largest eigenvalues are degenerate, the partition function (in the thermodynamic limit, with periodic boundary conditions) will vanish. As an example, with $z = 4$ and $r = 2.55$ we obtain $\psi = 0.376$, $\chi = 2.872$, and then $x = -3.53831 + 1.50236i$, $y = -3.83123 - 1.626732$. Clearly, $|y| > |x|$ (by some 8%). Of course, x and y could be interchanged.

5. DISCUSSION

We can only speculate that similar results would be obtained for a Widom–Rowlinson lattice gas in a higher dimension, but it now seems most likely. At least we know that the dimension-independent circle theorem[4] cannot be extended to low activities for any arbitrary lattice system. There may, however, be other more appropriate descriptions of the distribution of zeros at low activities that would be independent of dimensionality or lattice type.

ACKNOWLEDGMENT

Many helpful discussions with J.L. Lebowitz are gratefully acknowledged. It was he who originally suggested look-

ing explicitly at the one-dimensional version of the general system studied in Ref. 4.

[1]L.K. Runnels and J.L. Lebowitz, J. Math. Phys. **15**, 1712 (1974).
[2]J.L. Lebowitz and G. Gallavotti, J. Math. Phys. **12**, 1129 (1971).
[3]C.N. Yang and T.D. Lee, Phys. Rev. **87**, 404, 410 (1952).
[4]L.K. Runnels and J.L. Lebowitz (to be published).
[5]S. Katsura and M. Ohminami, J. Phys. Soc. Jpn. A **5**, 95 (1972).

Metastable states in the infinite Ising model

Paul Vanheuverzwijn[a), b)]
Department of Physics, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 19 June 1979; accepted for publication 9 August 1979)

Using a characterization of metastability recently introduced by Sewell, taking the "lifetime" of a state to be determined by the evolution of the polarization density only, we show that the infinite Ising model can support metastable states, describing polarization in opposition to a weak external field, even in the presence of globally acting thermal perturbations.

1. INTRODUCTION

Theories of metastability can be classified according to their applicability to systems with either long or short range forces, as well as according to their structural or dynamical nature. A structural characterization usually involves some notion of thermodynamical stability or the presence of an analytic continuation of equilibrium states, whereas dynamical theories may emphasize the long lifetime of the state, suitably defined. In ideal cases, it is believed, the structural properties ensure the dynamical ones.

Among the rigorous results obtained so far, we mention in particular the scheme proposed by Penrose and Lebowitz, offering a dynamical characterization for systems with both long[1] and short[2] range forces, a study by Griffiths, Weng, and Langer[3] on the slow relaxation of the polarization density in the Ising–Weiss model (see also Ref. 4), and finally the results of Lanford and Ruelle[5] concerning the absence of metastability in a finite range lattice gas, when a structural characterization is adhered to.

Recently, however, various aspects of metastability have been combined in a general scheme proposed by Sewell in Ref. 6. There it was proposed that one may distinguish "ideal" and "normal" metastability, the former, though realizable in systems with long range forces only, having such strong structural properties that they ensure the infinite lifetime (see e.g., Refs. 7–9) and the latter, while having much weaker stability properties, seemingly describing metastability as it is observed in most cases.

For classical lattice and continuous systems with hard cores,[7] as well as for quantum lattice systems,[9] it is possible to define metastability by three properties. As only a very limited class of model systems may support such states—in particular, short range systems are excluded—this metastability has been termed "ideal" in Ref. 6.

Definition 1.1[6]: A state ω is said to be ideally metastable if it satisfies the following three conditions:

1. It does not minimize the system's free energy density functional, i.e., it is not globally thermodynamically stable.

2. For all bounded regions, ω minimizes the restriction of the free energy density functional to the set of states differing from ω only in that region, i.e., ω is locally thermodynamically stable, and therefore[8] satisfies the KMS conditions, or DLR conditions in the classical case.

3. ω minimizes the restriction of the free energy density functional to some reduced state space $\Omega^{(0)}$ that is closed under space translations and convex combinations.

One readily sees that condition 2 in the above definition is too strong a property to be required for metastability in general systems. States satisfying the following weaker conditions will be called normal metastable states. In extreme cases the definition reduces to that of 1.1.

Definition 1.2[6]: Let $\mathscr{A}$ be the algebra of observables for the system, $\mathscr{A}_0$, a subset of $\mathscr{A}$, and Γ a class of dynamical semigroups of $\mathscr{A}$ (i.e., strongly continuous semigroups of completely positive maps γ_t ($t \in \mathbb{R}^+$),[10] arising from localized couplings of the system to some reservoir in thermal equilibrium. Let $\Omega^{(0)}$ be some reduced state space of $\mathscr{A}$ as in Definition 1.1. A state $\omega^{(0)}$ in $\Omega^{(0)}$ is then said to be metastable with respect to $(\Omega^{(0)}, \mathscr{A}_0, \Gamma)$ if, besides conditions 1 and 3 above, it has the property.

2′. There is a function ψ on $\mathbb{R}^+$ and a positive number τ, corresponding to a time that is very long by observational standards, such that, for all A in $\mathscr{A}_0$ and all γ in Γ,

$$|\omega^{(0)}(\gamma_t(A)) - \omega^{(0)}(A)| < \psi(t/\tau)\|A\|$$

with $\lim_{s \to 0} \psi(s) = 0$ and, by preference, $\tau > \tau_0 \times \exp(B/kT)$ with τ_0 of the order of standard observational times, and B an activation energy required to remove the system from $\Omega^{(0)}$.

Condition 2′ is a way of defining the lifetime of a state; it requires that, in the presence of certain globally acting thermal perturbations, the expectation values of a class of relevant observables decay "extremely slowly" to their equilibrium values. It is important to compare this definition with the one in Ref. 1 that measures lifetimes in terms of the conditional probability that the system being in $\Omega^{(0)}$ at time zero has escaped from $\Omega^{(0)}$ by time t. Roughly speaking, the set $\mathscr{A}_0$ in Definition 1.2 has been taken as the set of all observables.

In Ref. 2 a study was made of metastability in finite Ising model systems, in the spirit of this latter characterization. It was found that, although decay rates from $\Omega^{(0)}$ can be made very small, they nevertheless grow with volume. Therefore, metastability is not possible in the infinite Ising model, at least not in the class of states studied there.

It is the purpose of this paper to show that the infinite

a)Aspirant N.F.W.O. Belgium; on leave of absence from the University of Leuven.
b)Permanent address: Instituut voor Theoretische Fysika, Celestijnenlaan 200D, B3030 Leuven, Belgium.

 0022-2488/79/122665-06$01.00

size Ising model may support normal metastable states in the sense of Definition 1.2, even subject to global thermal perturbations. We shall consider states polarized in opposition to a weak external field, and thermal perturbations as described by the Glauber[11] dynamics. A rigorous proof that this semigroup actually arises from a coupling to a Markovian reservoir in equilibrium was given in Ref. 4, whereas an extention of the dynamics to an infinite system was shown to exist in Ref. 12. This model, commonly referred to as the stochastic Ising model, proved to be of great use in studying equilibrium properties.[13]

The reduced space of "wrongly polarized" states is idealized as follows.[2] Supposing the external field to be negative, we consider the state space $\Omega^{(c)}$, c being a positive number, consisting of those states that give zero weight to configurations where clusters of negative spin are present, with area exceeding a given value of c^2. It is hoped to single out *a posteriori* a critical value which best satisfies the metastability requirements.

We find that, for low temperatures and for c chosen in a suitable interval, any state minimizing the restriction of the free energy density functional to $\Omega^{(c)}$ has property 2′ and is therefore $(\Omega^{(c)}, \mathscr{A}_0, \Gamma)$-metastable if $\mathscr{A}_0$ is taken to be the set of spin observables and Γ the globally acting Glauber dynamics.

The article is organized as follows. In Sec. 2 we construct the relevant class of states and the dissipative dynamics. The basic estimates are derived in Sec. 3, the main result being Theorem 3.3. Finally, a brief discussion is presented in Sec. 4.

2. STATES AND DYNAMICS

As we shall consider states on the infinite Ising lattice, we first introduce some appropriate notation. $X = \{-1,1\}^{\mathbb{Z}^2}$ will denote the set of spin configurations on the lattice, whereas $X_\Lambda, X_{\tilde\Lambda}$ (with $\tilde\Lambda = \mathbb{Z}^2 \setminus \Lambda$) are taken to be the set of configurations in $\Lambda, \tilde\Lambda$, respectively. In general Λ will be a finite subset of $\mathbb{Z}^2$. Endowed with the product topology, all three spaces X, X_Λ, and $X_{\tilde\Lambda}$ are compact.

Given $x \in X$, x_Λ represents its projection on X_Λ. We shall occasionally write x as $(x_\Lambda, x_{\tilde\Lambda})$. Conversely, any x_Λ in X_Λ can always be considered as an element of X by identifying x_Λ with $(x_\Lambda, p_{\tilde\Lambda})$. Here $p_{\tilde\Lambda}$ denotes the configuration in $\tilde\Lambda$ in which all spins are positive. When no confusion is possible, we shall delete the subscript Λ and consider x as an element of either X_Λ or X.

Next, we define two transformations on X. If $x \in X$ and $j,k \in \mathbb{Z}^2$, then x_j will denote the configuration defined by

$$x_j(k) = x(k) \quad \text{for } j \neq k$$
$$= -x(k) \quad \text{for } j = k, \tag{2.1}$$

and x^j denotes the configuration

$$x^j(k) = x(k-j) \quad \text{for all } k. \tag{2.2}$$

It is particularly useful for our purposes to represent each configuration by a set of nonintersecting polygonal contours that separate spins of opposite sign (Ref. 14 and references therein). A set of contours, together with the specification of the spin value at a particular lattice point, uniquely determines the configuration. In this language an outer contour is one which is not enclosed by another contour. If ζ is any finite contour, then $|\zeta|$ is its length, $\theta(\zeta)$ the region enclosed by ζ, and $|\theta(\zeta)|$ the area of this latter set. We now introduce an infinite-volume version of a reduced configuration space used in Ref. 2.

Definition 2.1: If c is any positive number, X^c is the set of all configurations x such that if ζ is any outer contour associated with x, for which $x(k) = -1$ for some k in $\theta(\zeta)$ and adjacent to ζ, then $|\theta(\zeta)| \leqslant c^2$. We also define X^c_Λ by $X^c_\Lambda = \{x \in X_\Lambda \mid \exists\, y \in X^c \text{ such that } y_\Lambda = x\}$ and then define $X^c_{\tilde\Lambda}$ analogously. Let χ^c, resp. χ^c_Λ, resp. $\chi^c_{\tilde\Lambda}$, be the characteristic function of X^c in X, etc.

Since a sequence $x^{(n)}$ in X converges to x if and only if for all finite Λ there is an n_0 such that for $n > n_0$, $x^{(n)}_\Lambda = x_\Lambda$, it follows that X^c is a closed set of X, and that it is therefore compact in the relative topology.

Let then $\mathscr{A}$ be the C^*-algebra of all complex continuous functions on X with the uniform topology. The local observables for the bounded region Λ are the elements of $\mathscr{A}_\Lambda$

$$\mathscr{A}_\Lambda = \{f \in \mathscr{A} \mid f(x) = f(x_\Lambda)\}.$$

For a given c we also define $\mathscr{A}^c$,

$$\mathscr{A}^c = \{f \in \mathscr{A} \mid f(x) = 0 \text{ for } x \notin X^c\}.$$

we can define $\mathscr{A}^c_\Lambda, \mathscr{A}^c_{\tilde\Lambda}$, analogously. We here introduce two other subsets of $\mathscr{A}$ that will be used later. First we may consider $\mathscr{A}_0$; if s_k is the observable $s_k(x) = x(k)$,

$$\mathscr{A}_0 = \{s_k \mid k \in \mathbb{Z}^2\}, \tag{2.3}$$

a set which will appear in the characterization of metastability; second, we shall need the following auxiliary algebra $\mathscr{A}_F$,

$$\mathscr{A}_F = \{f \in \mathscr{A} \mid f(x) = f(x_\Lambda) \text{ for all } x \text{ and for some finite } \Lambda\} \tag{2.4}$$

There is a one-to-one correspondence between states on $\mathscr{A}$, and (regular Borel) probability measures on X. We shall call this latter set Ω. Similarly Ω^c is the set of (regular Borel) probability measures on X^c (isomorphic to the state space of $\mathscr{A}^c$). Ω_Λ and Ω^c_Λ then have an obvious meaning.

Definition 2.2: $\Omega^{(c)}$ is the convex set of (regular Borel) probability measures on X, with support in X^c. The map

$$\Omega^{(c)} \to \Omega^c$$
$$\omega^{(c)} \to \omega^c = \omega^{(c)}\big|_{X^c} \tag{2.5}$$

is an isomorphism.

We now introduce a particular class of states in $\Omega^{(c)}$; in Sec. 3 we shall show that at least one of these is a metastable state. The local Ising Hamiltonian in the presence of an external magnetic field h (that will be taken negative in the sequel) is defined as follows. If the system is in the configuration x, then its energy for the region Λ is given by

$$H_\Lambda(x) = -\tfrac{1}{2}h \sum_{k \in \Lambda} s_k(x) - J \sum_{k,k' \in \Lambda}{}' (s_k\, s_{k'})(x)$$
$$- J \sum_{\substack{k \in \Lambda \\ k' \in \tilde\Lambda}} (s_k\, s_{k'})(x). \tag{2.6}$$

Here the primed sum indicates summation over nearest neighbors, with each pair counted once only. We refer in a formal way to the "interaction $H(J,h)$."

Definition 2.3: We shall denote by $\omega^{(c)}_{\beta,h}$ any translationally invariant state that satisfies the DLR equations for the constraint system $(\mathscr{A}^c,\chi^c,\Omega^{(c)})$ with respect to the interaction $H(J,h)$ and at inverse temperature β.

In other words, $\omega^{(c)}_{\beta,h}$ satisfies the following set of equations.[15,16,5] For all finite Λ, $k\in\Lambda$, $x_\Lambda\in X_\Lambda$, and $x_{\tilde\Lambda}\in X_{\tilde\Lambda}$ (using Definition 2.1)

$$\chi^{(c)}((x_\Lambda,x_{\tilde\Lambda}))\omega^{(c)}_\Lambda[x_\Lambda;dx_{\tilde\Lambda}] = \omega^{(c)}_\Lambda[x_\Lambda;dx_{\tilde\Lambda}], \tag{2.7a}$$

$$\begin{aligned}\chi^{(c)}((x_\Lambda,x_{\tilde\Lambda}))&\omega^{(c)}_\Lambda[(x_\Lambda)_k;dx_{\tilde\Lambda}]\\ &= \chi^{(c)}(((x_\Lambda)_k,x_{\tilde\Lambda}))\exp[-\beta H_\Lambda((x_\Lambda)_k,x_{\tilde\Lambda})\\ &\quad+\beta H_\Lambda((x_\Lambda,x_{\tilde\Lambda}))]\omega^{(c)}_\Lambda[x_\Lambda;dx_{\tilde\Lambda}]\end{aligned} \tag{2.7b}$$

Here, $\omega^{(c)}_\Lambda[x_\Lambda;.]$ is the measure on $X_{\tilde\Lambda}$ for which

$$\omega^{(c)}(f) = \sum_{x_\Lambda\in X_\Lambda}\int\omega^{(c)}_\Lambda[x_\Lambda;dx_{\tilde\Lambda}]f((x_\Lambda,x_{\tilde\Lambda}))$$

for all f in $\mathscr{A}$. Clearly $\omega^{(c)}_{\beta,h}$ belongs to $\Omega^{(c)}$.

Adapting the techniques in Refs. 16 and 5, we obtain an existence theorem by using the correspondence (2.5). Let $a\in\mathbb{Z}^2$. Defining the map

$$\begin{aligned}&\tau_a:\mathscr{A}\to\mathscr{A},\\ &f\mapsto\tau_a f,\\ &(\tau_a f)(x) = f(x^a).\end{aligned} \tag{2.8}$$

we say that a sequence of probability measures μ_{Λ_n} on X_{Λ_n} (with Λ_n a sequence of finite volumes tending to infinity in the sense of Van Hove), averaged over translations, converges to a measure μ on X if for all f in $\mathscr{A}$ we have [16,5]

$$\lim_{n\to\infty}\frac{1}{|\Lambda_n|}\sum_{a\in\Lambda_n}\sum_{x\in X_{\Lambda_n}}\mu_{\Lambda_n}(x)(\tau_{-a}f)(x) = \mu(f), \tag{2.9}$$

In this case we write

$$\lim_{n\to\infty}\bar\mu_{\Lambda_n}(f) = \mu(f).$$

Theorem 2.4: Let Λ_n be as above. Let $\omega^c_{\Lambda_n}$ be the probability measure on Λ_n, assigning to each x in X_{Λ_n} the probability

$$\omega^c_{\Lambda_n}(x) = [Z^c_{\Lambda_n}]^{-1}\exp[-\beta H_\Lambda((x,p_{\tilde\Lambda}))]\chi^c_{\Lambda_n}(x) \tag{2.10}$$

with

$$Z^c_{\Lambda_n} = \sum_{y\in X_{\Lambda_n}}\exp[-\beta H_\Lambda((y,p_{\tilde\Lambda}))]\chi^c_{\Lambda_n}(y).$$

Then there is a subsequence $\Lambda_{n'}$ and a state $\omega^{(c)}_{\beta,h}$ as in Definition 2.3 such that, for all finite Λ and all $f\in\mathscr{A}_\Lambda$,

$$\lim_{n'\to\infty}\bar\omega^c_{\Lambda_{n'}|\Lambda}(f) = \omega^{(c)}_{\beta,h}(f)$$

Moreover, $\omega^{(c)}_{\beta,h}$ minimizes the restriction of the free energy density functional to $\Omega^{(c)}$. ■

So far, we have established the existence of at least one state, with the properties 1 and 3 in Definition 1.2.[17] That $\omega^{(c)}_{\beta,h}$ is not globally thermodynamically stable is seen by the fact that Eqs. (2.7) are incompatible with the DLR equations without constraints.

Let us now introduce a class of dynamical semigroups.[11-13] For a given $k\in\mathbb{Z}^2$,

define $\langle k\rangle$ as the set of nearest neighbors of k (2.11)

and $c(k,\cdot)$ as the observable

$$c(k,x) = \left[1+\exp\left[\beta hx(k)+2\beta Jx(k)\sum_{l\in\langle k\rangle}x(l)\right]\right]^{-1} \tag{2.12}$$

For fixed k and x, $c(k,x)$ is the (normalized) conditional probability of a spin-reversal at k, given the configuration x.

Theorem 2.5 (*Ref.* 12, *Theorem* 4.2): Define the map $\mathscr{L}_F:\mathscr{A}_F\to\mathscr{A}_F$ by

$$(\mathscr{L}_F f)(x) = \sum_{j\in\mathbb{Z}^2}c(j,x)[f(x_j)-f(x)]; \tag{2.13}$$

Then the closure $\mathscr{L}$ of $\mathscr{L}_F$ generates a dynamical semigroup γ_t, $t\in\mathbb{R}^+$, on $\mathscr{A}$. Moreover, if, given a finite Λ, $\mathscr{L}_\Lambda$ is the map $\mathscr{L}_\Lambda:\mathscr{A}_\Lambda\to\mathscr{A}_\Lambda$,

$$(\mathscr{L}_\Lambda f)(x) = \sum_{j\in\Lambda}c(j,x)[f(x_j)-f(x)] \tag{2.14}$$

and γ^Λ_t the semigroup it generates, then, for all g in $\mathscr{A}_\Lambda$ and all finite t_0,

$$\lim_{\Lambda\to\infty}\sup_{0\le t\le t_0}\|\gamma^\Lambda_t(g)-\gamma_t(g)\| = 0.$$ ■

We may consider τ_a, (2.8), as a map $\mathscr{A}_\Lambda\to\mathscr{A}_{\Lambda-a}$ and obtain

Lemma 2.6: (1) For all $a\in\mathbb{Z}^2$, for all finite Λ: $\tau_a\gamma^\Lambda_t = \gamma^{\Lambda-a}_t\tau_a$

(2) γ satisfies the principle of detailed balance.

Proof: (1) Let f be in $\mathscr{A}_\Lambda$ and x in X. Then

$$\begin{aligned}[(\tau_a\mathscr{L}_\Lambda)f](x)&\\ &= \sum_{j\in\Lambda}c(j,x^a)[f((x^a)_j)-f(x^a)]\\ &= \sum_{j\in\Lambda}c(j-a,x)[f((x^a)_j)-f(x^a)]\\ &= \sum_{j\in\Lambda-a}c(j,x)[f((x^a)_{j+a})-f(x^a)]\\ &= \sum_{j\in\Lambda-a}c(j,x)[f((x_j)^a)-f(x^a)].\end{aligned}$$

(2) This principle, which asserts that, for any finite Λ, any $j\in\Lambda$, and any $x\in X$,

$$c(j,x)\exp[-\beta H_\Lambda(x)] = c(j,x_j)\exp[-\beta H_\Lambda(x_j)],$$

is easily verified using (2.6) and (2.13). ■

As is well known, the property of detailed balance is the result of a coupling to a Markovian reservior. See Ref. 4 for a derivation of a (more general) Glauber-type semigroup from first principles.

Lemma 2.7: For all f in $\mathscr{A}$, for all $h<0$, and for all β,

$$\lim_{t\to\infty}\omega^{(c)}_{\beta,h}(\gamma_t(f)) = \omega_{\beta,h}(f).$$

Here $\omega_{\beta,h}$ is the (unique) equilibrium state for the interaction $H(J,h)$, at inverse temperature β.

Proof: This is a direct consequence of Ref. 19, Appendix, and Corollary 3.17 in Ref. 13. ■

3. METASTABILITY

As yet, no conditions have had to be imposed in order to show the existence of a state $\omega^{(c)}_{\beta,h}$, having properties 1 and 3 of Definition 1.2. We shall finally obtain an estimate on the dynamics of the polarization densities so as to establish property 2'. Since we shall rely partially on an estimate in Ref. 2, we must now impose the following conditions on β and c:

$$\beta > 2(\ln 6)/J, \tag{3.1}$$

$$c < 4(\beta J - 2\ln 6)/\beta |h|. \tag{3.2}$$

We shall temporarily fix β and h in what follows, and thus delete all subscripts referring to these parameters. Let us denote the set

$$\{x \in X_\Lambda \,|\, x \in X^c_\Lambda; x_j \notin X^c_\Lambda\} \quad \text{by} \quad X^c_\Lambda(j). \tag{3.3}$$

Lemma 3.1: Let Λ, $\bar\Lambda$ be the two finite subsets of $\mathbb{Z}^2$, with k a fixed lattice point in $\bar\Lambda$. Take ω^c_Λ as in Theorem 2.4, γ^Λ_t as in Theorem 2.5. If Λ' is the set $\{a \in \mathbb{Z}^2 | a + \bar\Lambda \subseteq \Lambda\}$, Then

$$\frac{d}{dt}\bar\omega^c_\Lambda[\gamma^{\bar\Lambda}_t(s_k)] = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{j\in\Lambda}\sum_{x\in X^c_\Lambda(j)} c(j,x)\cdot\frac{1}{Z^c_\Lambda} \times[(\gamma^{\bar\Lambda+a}_t s_{k+a})(x_j)] - [\gamma^{\bar\Lambda+a}_t s_{k+a})(x)] \times\exp[-\beta H_\Lambda(x)] \tag{3.4}$$

Proof: We shall identify x with $(x, p_{\bar\Lambda})$. We obtain, by (2.9) and Lemma 2.6,

$$\bar\omega^c_\Lambda(\gamma^{\bar\Lambda}_t s_k) = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{x\in X_\Lambda}\omega^c_\Lambda(x)(\tau_{-a}\gamma^{\bar\Lambda}_t s_k)(x) = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{x\in X_\Lambda}\omega^c_\Lambda(x)(\gamma^{\bar\Lambda+a}_t\tau_{-a}s_k)(x) = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{x\in X_\Lambda}\omega^c_\Lambda(x)(\gamma^{\bar\Lambda+a}_t s_{k+a})(x).$$

Therefore, by (2.14) and using the fact that $\gamma^{\bar\Lambda+a}_t s_{k+a}$ is independent of the coordinates not in $(a+\bar\Lambda)$, we have

$$\frac{d}{dt}\bar\omega^c_\Lambda(\gamma^{\bar\Lambda}_t s_k) = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{x\in X_\Lambda}\omega^c_\Lambda(x)\sum_{j\in\Lambda}c(j,x) \times[(\gamma^{\bar\Lambda+a}_t s_{k+a})(x_j) - (\gamma^{\bar\Lambda+a}_t s_{k+a})(x)] = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{j\in\Lambda}\sum_{x\in X_\Lambda}(\gamma^{\bar\Lambda+a}_t s_{k+a})(x) \times[c(j,x_j)\omega^c_\Lambda(x_j) - c(j,x)\omega^c_\Lambda(x)] = \frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{j\in\Lambda}(\gamma^{\bar\Lambda+a}_t s_{k+a})(x)\cdot\frac{1}{Z^c_\Lambda} \times\Bigg(\sum_{\substack{\{x\in X_\Lambda|\\ x_j\in X^c_\Lambda\}}} c(j,x_j)\exp[-\beta H_\Lambda(x_j)] - \sum_{\substack{\{x\in X_\Lambda|\\ x\in X^c_\Lambda\}}} c(j,x)\exp[-\beta H_\Lambda(x)]\Bigg). \tag{3.5}$$

Applying Lemma 2.6 again, (3.5) equals

$$= \frac{1}{Z^c_\Lambda}\frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{j\in\Lambda}\Bigg(\sum_{\substack{\{x\in X_\Lambda|\\ x_j\in X^c_\Lambda\}}}(\gamma^{\bar\Lambda+a}_t s_{k+a})(x) \times c(j,x)\exp[-\beta H_\Lambda(x)] - \sum_{\substack{\{x\in X_\Lambda|\\ x\in X^c_\Lambda\}}}(\gamma^{\bar\Lambda+a}_t s_{k+a})(x) \times c(j,x)\exp[-\beta H_\Lambda(x)]\Bigg)$$

$$= \frac{1}{Z^c_\Lambda}\frac{1}{|\Lambda'|}\sum_{a\in\Lambda'}\sum_{j\in\Lambda}\Bigg(\sum_{\substack{\{x\in X_\Lambda|\\ x_j\in X^c_\Lambda(j)\}}}(\gamma^{\bar\Lambda+a}_t s_{k+a})(x) \times c(j,x)\exp[-\beta H_\Lambda(x)] - \sum_{x\in X^c_\Lambda(j)}(\gamma^{\bar\Lambda+a}_t s_{k+a})(x)c(j,x)\exp[-\beta H_\Lambda(x)]\Bigg)$$

A "change of variables" in the first sum ($x \mapsto x_j$) and a final application of Lemma 2.6 then lead to (3.4). ■

So far, not very much information about the dynamics has been used, apart from detailed balance and translation invariance. The finite range structure of the dynamics, however, gives us a bound on the propagation velocity of Γ, which is volume independent, even if contributions of all lattice sites are added. If j and k are two lattice points, then $|j-k|$ denotes their distance.

Lemma 3.2: If k, Λ, $\bar\Lambda$ are chosen as in Lemma 3.1, if $\kappa = \sup_{j\in\mathbb{Z}^2}\sup_{x\in X}|c(j,x)|$ and $\nu = |\langle k\rangle|$, the number of nearest neighbors of k, then

$$\sum_{a\in\Lambda'}|(\gamma^{\bar\Lambda+a}_t s_{k+a})(x_j) - (\gamma^{\bar\Lambda+a}_t s_{k+a})(x)| \leqslant 2\nu^{-1}[\exp(2\kappa\nu t) - 1] + 2 + 32\kappa t\exp(4\kappa\nu t) \equiv g(t). \tag{3.6}$$

Proof:

$$(\gamma^{\bar\Lambda+a}_t s_{k+a})(x_j) - (\gamma^{\bar\Lambda+a}_t s_{k+a})(x) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\{[(\mathscr{L}_{\bar\Lambda+a})^n s_{k+a}](x_j) - [(\mathscr{L}_{\bar\Lambda+a})^n s_{k+a}](x)\}$$

For general Λ and k we have, for $n \geqslant 1$,

$$(\mathscr{L}^n_\Lambda s_k)(x_j) - (\mathscr{L}^n_\Lambda s_k)(x) = \sum_{(l_2,\ldots,l_n)\in\Lambda^{n-1}}\sigma^{(n)}_j(l_2,\ldots,l_n;x) \tag{3.7}$$

The term $\sigma^{(n)}_j(l_2,\ldots,l_n;x)$ is k-dependent; as we temporarily keep k fixed, we shall omit any reference to it. The $n=1$ term will be denoted by $\sigma^{(1)}_j(x)$. Formula (3.7) should be read as follows:

$$\sigma^{(1)}_j(x) = -2c(k,x_j)s_k(x_j) + 2c(k,x)s_k(x), \tag{3.8}$$

$$\sigma^{(n)}_j(l_2,\ldots,l_n;x) = c(l_n,x_j)\sigma^{(n-1)}_{l_n}(l_2,\ldots,l_{n-1};x_j) - c(l_n,x)\sigma^{(n-1)}_{l_n}(l_2,\ldots,l_{n-1};x). \tag{3.9}$$

These recursion expressions are readily obtained using (2.14) and the fact that $s_k(x)$ depends on the k th coordinate only.

We show that, for all x and all $(l_2,\ldots,l_n)\in\Lambda^{n-2}$,

$$|k-j| > n \Rightarrow \sigma^{(n)}_j(l_2,\ldots,l_n;x) = 0. \tag{3.10}$$

Clearly the result is true for $n = 1$, as is seen by inspection of the form (3.8). Suppose now that the result (3.10) is true for $n-1$, and suppose $|j-k|>n$.

(a) Let $|l_n - k|>n-1$, then, by the induction hypothesis, it follows that

$$\sigma_{l_n}^{(n-1)}(l_2,\ldots,l_{n-1};x_j)=0$$

as well as

$$\sigma_{l_n}^{(n-1)}(l_2,\ldots,l_{n-1};x)=0.$$

Therefore, using (3.9), we obtain $\sigma_j^{(n)}(l_2,\ldots,l_n;x)=0$.

(b) Let $l_n\notin\langle j\rangle$, then we obtain, using (3.9) and introducing, when necessary, $\sigma_j^{(0)}(x)=s_k(x_j)-s_k(x)$,

$$\begin{aligned}
&\sigma_j^{(n)}(l_2,\ldots,l_n;x)\\
&\quad=c(l_n,x)\left[\sigma_{l_n}^{(n-1)}(l_2,\ldots,l_{n-1};x_j)\right.\\
&\qquad\left.-\sigma_{l_n}^{(n-1)}(l_2,\ldots,l_{n-1};x)\right]\\
&\quad=c(l_n,x)\left[c(l_{n-1},x_{jl_n})\right]\sigma_{l_n l_{n-1}}^{(n-2)}(l_2,\ldots,l_{n-2};x_{jl_n})\\
&\qquad-c(l_{n-1},x_j)\sigma_{l_{n-1}}^{(n-2)}(l_2,\ldots,l_{n-2};x_j)\\
&\qquad-c(l_{n-1},x_{l_n})\sigma_{l_{n-1}}^{(n-2)}(l_2,\ldots,l_{n-2};x_{l_n})\\
&\qquad+c(l_{n-1},x)\sigma_{l_{n-1}}^{(n-2)}(l_2,\ldots,l_{n-2};x)\\
&\quad=c(l_n,x)\left[\sigma_j^{(n-1)}(l_2,\ldots,l_{n-1};x_{l_n})\right.\\
&\qquad\left.-\sigma_j^{(n-1)}(l_2,\ldots,l_{n-1};x)\right],
\end{aligned}$$

which is zero, because of the induction hypothesis.

As it is impossible, when $|j-k|>n$, to find l_n such that $|l_n-k|\leqslant n-1$ and simultaneously $l_n\in\langle j\rangle$, we conclude that

$$\sigma_j^{(n)}(l_2,\ldots,l_n;x)=0 \quad \text{for } |j-k|>n.$$

Note also that

$$\sigma_j^{(0)}(x)=0 \quad \text{whenever } j\neq k.$$

If, on the other hand, $|k-j|\leqslant n$, there is possibly a contribution to the nth term, which, however, by (3.9) and the above analysis is bounded by

$$\begin{aligned}
&2 \text{ for } n=0,\\
&(t^n/n!)\kappa^n\nu^{n-1}2^{n+1} \quad \text{for } n\geqslant 1.
\end{aligned}\tag{3.11}$$

Fixing j, we finally have to add up contributions from all lattice points k. We subdivide the lattice in a number of squares, the zeroth consisting of $\{j\}$ only, the nth having vertices $j+(\pm n,\pm n)$, and thus consisting of $8n$ points, a distance $d\geqslant n$ away from j.

Because of the results (3.10) and (3.11), we know that the contribution of the nth square ($n\geqslant 1$) is bounded by

$$8n\sum_{m=n}^{\infty}\frac{t^m}{m!}\kappa^m\nu^{m-1}2^{m+1}$$

hence

$$\begin{aligned}
&\sum_{a\in n'}\left|(\gamma_t^{\bar\Lambda+{}^a s_{k+a}})(x_j)-(\gamma_t^{\bar\Lambda+{}^a s_{k+a}})(x)\right|\\
&\quad\leqslant 2+\sum_{n=1}^{\infty}\frac{t^n}{n!}\kappa^n\nu^{n-1}2^{n+1}\\
&\qquad+\sum_{n=1}^{\infty}8n\sum_{m=n}^{\infty}\frac{t^m}{m!}\kappa^m\nu^{m-1}2^{m+1}\\
&\quad\leqslant 2+2\nu^{-1}[\exp(2\nu\kappa t)-1]\\
&\qquad+16\nu^{-1}\sum_{n=1}^{\infty}n\sum_{m=n}^{\infty}\frac{(2\nu\kappa t)^m}{m!}\\
&\quad\leqslant 2+2\nu^{-1}[\exp(2\nu\kappa t)-1]\\
&\qquad+16\nu^{-1}\sum_{n=1}^{\infty}n\,[\exp(2\nu\kappa t)]\frac{(2\nu\kappa t)^n}{n!}.
\end{aligned}$$

■

Although the bound (3.6) is Λ independent, it is presumably not the best possible, since it grows exponentially with t. The occurrence of a summation over the lattice, however, seriously worsens the situation.

We now state the main result, which, after a suitable range of values for β and c has been chosen, will entail the existence of a metastable state with respect to $(\Omega^{(c)},\mathscr{A}_0,\Gamma)$.

Theorem 3.3: If $\beta>2\ln 6/J$ and $c<4(\beta J-2\ln 6)/\beta|h|$, $h<0$, there exists a state $\omega_{\beta,h}^{(c)}$, as in Definition 2.3, and a function G of (t,β,h,c,κ), bounded in β, h, and c such that

$$\begin{aligned}
&\left|\omega_{\beta,h}^{(c)}[\gamma_t(s_k)-s_k]\right|\leqslant G(t,\beta,h,c,\kappa)\\
&\qquad\times\exp[-(\beta J-\ln 3)4c+\beta|h|c^2].
\end{aligned}\tag{3.12}$$

Proof: It follows from Lemma 3.1 that

$$\begin{aligned}
&\left|\bar\omega_\Lambda^c[\gamma_t^{\bar\Lambda}(s_k)-s_k]\right|\\
&\quad=\frac{1}{|\Lambda'|}\left|\sum_{a\in\Lambda'}\sum_{j\in\Lambda}\sum_{x\in X_\Lambda^c(j)}c(j,x)\right.\\
&\qquad\times\exp[-\beta H_\Lambda(x)]\cdot[Z_\Lambda^c]^{-1}\\
&\qquad\left.\times\int_0^t du[(\gamma_u^{\bar\Lambda+{}^a s_{k+a}})(x_j)-(\gamma_u^{\bar\Lambda+{}^a s_{k+a}})(x)]\right|.
\end{aligned}\tag{3.13}$$

Hence, by Lemma 3.2, letting $i(t)$ be $\int_0^t du\,g(u)$, we obtain the upper bound for (3.13):

$$\leqslant i(t)\frac{1}{|\Lambda'|}\sum_{j\in\Lambda}\sum_{x\in X_\Lambda^c(j)}c(j,x)\exp[-\beta H_\Lambda(x)]\cdot[Z_\Lambda^c]^{-1}$$

If now conditions (3.1) and (3.2) are satisfied, then by Ref. 2, Theorem 2, there is a bounded function F such that the above, in turn, is bounded by

$$\begin{aligned}
&\leqslant i(t)F(\beta,h,c)(|\Lambda|/|\Lambda'|)\kappa\\
&\qquad\times\exp[-(\beta J-\ln 3)4c+\beta|h|c^2]\\
&\equiv G(t,\beta,h,c,\kappa)(|\Lambda|/|\Lambda'|)\\
&\qquad\times\exp[-(\beta J-\ln 3)4c+\beta|h|c^2]
\end{aligned}$$

Finally, let $\omega_{\beta,h}^{(c)}$ be any state as constructed in Theorem 2.4, Λ a typical element in the approximating subsequence, and $\bar\Lambda$ such that $k\in\bar\Lambda$. Then, deleting subscripts, we have

$$\begin{aligned}
&\left|\omega^{(c)}[\gamma_t(s_k)-s_k]\right|\\
&\quad\leqslant\left|\omega^{(c)}[\gamma_t(s_k)-s_k]-\omega^{(c)}[\gamma_t^{\bar\Lambda}(s_k)-s_k]\right|\\
&\qquad+\left|\omega^{(c)}[\gamma_t^{\bar\Lambda}(s_k)-s_k]-\bar\omega_\Lambda^c[\gamma_t^{\bar\Lambda}(s_k)-s_k]\right|\\
&\qquad+\left|\bar\omega_\Lambda^c[\gamma_t^{\bar\Lambda}(s_k)-s_k]\right|\\
&\quad\leqslant\|\gamma_t(s_k)-\gamma_t^{\bar\Lambda}(s_k)\|\\
&\qquad+\left|(\omega^{(c)}-\bar\omega_\Lambda^c)(\gamma_t^{\bar\Lambda}(s_k)-s_k)\right|\\
&\qquad+G(t,\beta,h,c,\kappa)(|\Lambda|/|\Lambda'|)\\
&\qquad\times\exp[-(\beta J-\ln 3)4c+\beta|h|c^2].
\end{aligned}$$

By Theorem 2.5, for all ϵ, there is a $\bar\Lambda_0$ such that, for $\bar\Lambda\supseteq\bar\Lambda_0$, $\|\gamma_t(s_k)-\gamma_t^{\bar\Lambda}(s_k)\|<\epsilon/2$. By hypothesis for all ϵ, for all $\bar\Lambda$,

there is a Λ_0, such that for all $\Lambda \supseteq \Lambda_0$

$$\left|(\omega^{(c)} - \bar{\omega}_c^{\Lambda})[\gamma_t^{\bar{\Lambda}}(s_k) - s_k]\right| < \epsilon/2$$

and

$$|\Lambda|/|\Lambda'| < 1 + \epsilon$$

as Λ tends to infinity in the sense of Van Hove. Hence, for all ϵ, we have

$$\left|\omega_{\beta,h}^{(c)}[\gamma_t(s_k) - s_k]\right| < \epsilon + (1+\epsilon)G(t,\beta,h,c,\kappa) \times \exp[-(\beta J - \ln 3)4c + \beta|h|c^2],$$

which proves the result. ■

We can now fix a range of values for c such that the metastability requirements are best satisfied, i.e., such that the lifetime of $\omega_{\beta,h}^{(c)}$ be as long as possible. For large β, it is shown in [2] that a maximum is reached if c belongs to the interval

$$\frac{2J}{|h|} - \left(\frac{2(J-|h|)}{|h|}\right)^{1/2} \leqslant c \leqslant \frac{2J}{|h|} + \left(\frac{2(J-|h|)}{|h|}\right)^{1/2},$$

in which case the bound turns out to be of the following nature:

$$\left|\omega_{\beta,h}^{(c)}[\gamma_t(s_k) - s_k]\right| < i(t)\kappa F(\beta,h,c) \exp[-4\beta J]^2/|h|$$

Upon calling $\tau(h) = \exp[4\beta J^2/|h|]$ (basically $J^2/|h|$ is the energy required to remove the system from $\Omega^{(c)}$), we obtain

$$\left|\omega_{\beta,h}^{(c)}[\gamma_t(s_k) - s_k]\right| < \psi(t/\tau(h)).$$

And in particular,

$$\lim_{h \to 0}\left|\omega_{\beta,h}^{(c)}[\gamma_t(s_k) - s_k]\right| = 0$$

i.e., the metastability becomes perfect in the limit $h \to 0$. For higher temperatures it can similarly be seen that the argument of the exponential in (3.12) is a bound on the activation free energy required to remove the system from $\Omega^{(c)}$.

4. DISCUSSION

As we stated earlier, the main point of this article is that we have shown that an infinitely extended (classical) system with short range forces can support metastable states when an appropriate characterization (Ref. 6, [4.2]) is adhered to.

We have constructed a state which, even in the presence of a globally acting thermal perturbation, has a very long lifetime, in the sense that the expectation values of the local spin observables remain constant in the limiting regime where the external magnetic field tends to zero. A criterion proposed in Ref. 1, precludes the possibility of metastability in the same infinite model,[2] as the results on the escape rate obtained there for finite systems indicate a proportionality to the volume.

It is of interest to note that computer simulations (Ref. 20 and references therein) already showed an absence of any volume dependence in the relaxation time of the spin observables in finite stochastic Ising models. This result was obtained using periodic boundary conditions. If on the other hand, open boundary conditions were used, then some dependence becomes apparent for small systems. We suspect, however, that for larger systems this dependence is not present, as it is a surface effect only. Moreover, at low temperatures, there seemingly is a unique metastable state.

We finally comment that the methods used here can be extended to any finite-range stochastic Ising model and to a class of semigroups with $c(k,x)$ as in Ref. 4. Indeed, the only difference is to multiply the $c(k,x)$ considered here with some function $g(k,x)$, depending on $[-\frac{1}{2}hx(k) - Jx(k)\Sigma_{l \in (k)} x(l)]$ only.

ACKNOWLEDGMENTS

This research was carried out at the suggestion of Dr. G.L. Sewell. It is a great pleasure to thank him for his warm hospitality at Queen Mary College, as well as for numerous suggestions and helpful advise during the course of preparation. Financial assistance both from the Belgian N.F.W.O. and the Instituut voor Theoretische Fysika, Leuven, are also gratefully acknowledged.

[1] J.L. Lebowitz and O. Penrose, J. Stat. Phys. **3**, 211–36 (1971).
[2] D. Capocaccia, M. Cassandro, and E. Olivieri, Comm. Math. Phys. **39**, 189–205 (1974).
[3] R. Griffiths, C. Weng, and J. Langer, Phys. Rev. **149**, 301 (1966).
[4] P. Martin, J. Stat. Phys. **16**, 149–68 (1977).
[5] O. Lanford and D. Ruelle, Comm. Math. Phys. **13**, 194–215 (1969).
[6] G.L. Sewell, "Stability, equilibrium, and metastability in statistical mechanics," preprint (1979).
[7] G.L. Sewell, Ann. Phys. **97**, 55–79 (1976).
[8] G.L. Sewell, Comm. Math. Phys. **55**, 53–61 (1977).
[9] G.L. Sewell, Comm. Math. Phys. **55**, 63–66 (1977).
[10] L. Accardi and V. Gorini (Eds.), *Mathematical Problems in the Theory of Quantum Irreversible Processes*, Proceedings, 1978 Naples Conference (to appear).
[11] R.J. Glauber, J. Math. Phys. **4**, 294–307 (1963).
[12] T.M. Liggett, Trans. Amer. Math. Soc. **165**, 471–81 (1972).
[13] R. Holley, Comm. Math. Phys. **23**, 87–99 (1971).
[14] G. Gallavotti, A. Martin-Löf, and S. Miracle-Sole, in *Statistical Mechanics and Mathematical Problems*, edited by A. Lenard, Lecture Notes in Physics **20** (Springer, Berlin, Heidelberg, New York 1973).
[15] C. Gruber and A. Hintermann, Physica A **83**, 233–56 (1976).
[16] O. Lanford, in Ref. 14.
[17] It seems reasonable that there is a β_1 which is c-independent, at least in a suitable range, such that for all $\beta > \beta_1$, there is a unique state $\omega_{\beta,h}^{(c)}$ for all c in that range and for all $h < 0$ (compare Corrollary to Theorem 4 in Ref. 18). In particular, there is a unique state $\omega_{\beta,h}^{(c)}$ at all temperatures for both $c = 0$ and $c = +\infty$, whenever $h \neq 0$.
[18] A. Hintermann and C. Gruber, Physica A **84**, 101–28 (1976).
[19] D. Ruelle, Ann. Phys. (N.Y.) **69**, 364–74 (1972).
[20] K. Binder and H. Muller-Krumbhaar, Phys. Rev. B **9**, 2328–53 (1974).

Continuum theory of nematic liquid crystals subject to electromagnetic fields[a)]

A. Cemal Eringen
Princeton University, Princeton, New Jersey 08540

(Received 10 August 1978; accepted for publication 26 October 1978)

A continuum theory is presented for the nematic liquid crystals subject to electromagnetic interactions. Complete field equations, jump conditions, and constitutive equations are obtained. The thermodynamic restrictions are studied. The field equations obtained are employed to examine the problem of Fredericksz' transition.

1. INTRODUCTION

In a previous paper[1] we gave a continuum theory of liquid crystals based on the micropolar continuum mechanics. Basic to this theory is the concept of orientable points with microinertia and degrees of freedom involving translations and rotations. As discussed, the liquid crystal theory so constructed contains the director theories in the special case of incompressibility, perfect alignment, and threadlike elements. An important consequence of the theory is that the liquid crystal theory rightly belongs to micropolar continuum mechanics. The present work represents further development of the liquid crystal theory in that the electromagnetic interactions with the flow field are taken into account.

Literature contains a number of papers dealing with the magnetic effect in liquid crystals. However most of this work is concerned with special situations and problems, e.g., effects of *uniform* magnetic field with flow orientations. The literature on electrical effects are even more scarce. The state of art up to 1974 in this field is well described in the review article by Stephen and Straley[2] and in deGennes' book.[3]

The purpose of the present article is to develop systematically the field equations, jump conditions, and constitutive equations of nematic liquid crystals from the basic principles of micropolar continuum theory. We make no assumption as to the nature of E–M fields and consider interactions of nonuniform electric field, magnetic field, heat conduction, and flow field in an orientable flow field. The elements of nematic liquid crystals need not be threadlike or incompressible and considered to possess arbitrary microinertia tensor. In Sec. 2 we summarize pertinent kinematical results. In Sec. 3 balance laws of micropolar continuum mechanics and electromagnetism are presented. The electromagnetic force and couple are those given by Dixon and Eringen,[4] DeGroot and Suttorp,[5] and Maugin and Eringen[6] in connection with deformable bodies. It was necessary to express Maxwell's equations in a comoving frame of reference with the points of the body. This is done (cf. Refs. 4–6) by means of Lorentz invariance of Maxwell's equations to within an approximation $1/c^2$ where c is the speed of light in vacuum. (For a fully Laurentz invariant theory see Grot and Eringen.[7])In Sec. 4, we develop the constitutive theory and study thermodynamic restrictions and the requirement of the material frame-indifference on the constitutive equations. Section 5 is devoted to the development of properly invariant constitutive equations for stress, couple stress, polarization, magnetization, heat, and electric conductions. Thermodynamic restrictions and the material stability on the constitutive moduli are discussed in Sec. 6. In Sec. 7 we indicate the passage to the director theory when the molecular elements of liquid crystal can be considered straight, threadlike filaments. The theory is now complete and ready for application. As an example, in Sec. 8, we give the solution of the classical Frederiksz' problem of rotation of nematic elements under a uniform magnetic field perpendicular to the long axes of molecules. The results of this solution are specialized to threadlike elements in Sec. 9. Finally computation has been carried out for this case and the results are plotted in Fig. 2. The results are gratifying in that they agree with what is already known experimentally.

2. KINEMATICS

In our previous formulation of mechanics of liquid crystals,[1] we have shown that the foundation of the theory of liquid crystals may be based upon the micropolar continuum mechanics. In a micropolar continuum a material point possesses mass density ρ and an inertia tensor j_{kl}. The motion of a material point is described fully by a translation and a rotation. Referred to a rectangular frame of reference X_K, $K = 1, 2, 3$ in the *reference state*, a material point $\mathbf{X}$ is characerized by its position vector X_K and a director Ξ attached to the point. The motion of $\mathbf{X}$ is then described by the following two sets of equations:

$$x_k = x_k(\mathbf{X},t), \quad \xi_k = \chi_{kK}(\mathbf{X},t)\Xi_K \tag{2.1}$$

of which the first one expresses the translatory motion of $\mathbf{X}$ and the second one orientation ξ_k of Ξ_K, at time t (the rotation of a microelement). The repeated indices are summed over 1, 2, 3. The inverse motions are posited to be unique, i.e.,

$$X_K = X_K(\mathbf{x},t), \quad \Xi_K = \chi^{-1}_{Kk}\xi_k, \tag{2.2}$$

where $\chi^{-1}_{Kk} = \chi_{kK}$, so that

$$\begin{aligned} &x_{k,K}X_{K,l} = \delta_{kl}, \quad X_{K,k}X_{k,L} = \delta_{KL}, \\ &\chi_{kK}\chi_{lK} = \delta_{kl}, \quad \chi_{kL}\chi_{kK} = \delta_{KL}, \end{aligned} \tag{2.3}$$

where an index followed by a comma indicates partial differentiation, e.g., $x_{k,K} \equiv \partial x_k/\partial X_K$, $X_{K,k} = \partial X_K/\partial x_k$. Henceforth we also employ a superposed dot to indicate the materi-

[a)]The present work was supported by the National Science Foundation.

 0022-2488/79/122671-11$01.00

al time derivative. The angular velocity vector ν_k is then introduced by

$$\dot{\xi} = \mathbf{v} \times \xi. \tag{2.4}$$

For a fluent body by considering the relative motions from the configurations at time t we have obtained strain measures, cf. Ref. 1,

$$\mathfrak{S}_{kl} = X_{K,k}\chi_{lK}, \quad \gamma_{kl} = \tfrac{1}{2}\epsilon_{kmn}\chi_{mK}\chi_{nK,l} \tag{2.5}$$

and the rate of deformation measures

$$a_{kl} = v_{l,k} + \nu_{kl}, \quad b_{kl} = \nu_{k,l} \tag{2.6}$$

where

$$\nu_{kl} = -\epsilon_{klm}\nu_m = \dot{\chi}_{kK}\chi_{lK} = -\nu_{lk} \tag{2.7}$$

is the so-called *gyration* tensor, where ϵ_{klm} is the alternating tensor. Indices following a comma denote partial differentiations with respect to space variables and a superposed dot indicates the material derivative.

We have shown that χ_{kK}, γ_{kl} and ν_k can be expressed in terms of an axial vector ϕ_k and its material time rate by (cf. Ref. 9, p. 20, 29, 34)

$$\chi_{kK} = [\cos\phi\delta_{kl} + (1-\cos\phi)n_k n_l - \sin\phi\epsilon_{klm}n_m]\,\delta_{lK}, \tag{2.8}$$

$$\gamma_{kl} = n_k\phi_{,l} + \sin\phi n_{k,l} - (1-\cos\phi)\,\epsilon_{kmn}n_m n_{n,l}, \tag{2.9}$$

$$\nu_k = \Lambda_{kl}\dot{\phi}_l, \tag{2.10}$$

where

$$n_k = \phi_k/\phi, \quad \phi = (\phi_k\phi_k)^{1/2}, \tag{2.11}$$

$$\Lambda_{kl} = \frac{\sin\phi}{\phi}\delta_{kl} + \left(1 - \frac{\sin\phi}{\phi}\right)n_k n_l - \frac{1}{\phi}(1-\cos\phi)\,\epsilon_{klm}n_m. \tag{2.12}$$

Here ϕ and $\mathbf{n}$ have simple physical interpretations: ϕ is the angle of rotation and $\mathbf{n}$ is the unit vector along the axis of rotation of a microelement. δ_{kl} is the Kronecker delta and δ_{kK} is the shifter (direction cosines between X_K and x_k). When the spatial and material frames of reference coincide then δ_{kK} is the same as the Kronecker delta.

3. BALANCE LAWS

The balance laws of liquid crystals are the same as those of the micropolar continua. Within the volume $\mathscr{V}$ of the body excluding the points that lie on a discontinuity surface σ which may be sweeping $\mathscr{V}$ with a velocity $\mathbf{u}$, these laws are expressed as (cf. Eringen[8,9])

Mass:

$$\frac{\partial\rho}{\partial t} + (\rho v_k)_{,k} = 0, \quad \mathscr{V} - \sigma, \tag{3.1}$$

Microinertia:

$$\frac{Dj_{kl}}{Dt} - \nu_{km}j_{lm} - \nu_{lm}j_{km} = 0, \quad \mathscr{V} - \sigma, \tag{3.2}$$

Momentum:

$$t_{kl,k} + \rho(f_l - \dot{v}_l) = 0, \quad \mathscr{V} - \sigma, \tag{3.3}$$

Moment of momentum:

$$m_{kl,k} + \epsilon_{lmn}t_{mn} + \rho(l_l - \dot{\sigma}_l) = 0, \quad \mathscr{V} - \sigma, \tag{3.4}$$

Energy:

$$\rho\dot{\epsilon} - t_{kl}a_{kl} - m_{kl}b_{lk} - q_{k,k} - \rho h = 0, \quad \mathscr{V} - \sigma, \tag{3.5}$$

Entropy inequality:

$$\rho\dot{\eta} - (q_k/\theta)_{,k} - \rho b \geqslant 0, \quad \mathscr{V} - \sigma, \tag{3.6}$$

where

$\rho =$ mass density,	$v_k =$ velocity vector,
$j_{kl} =$ microinertia tensor,	$\nu_{kl} =$ gyration tensor,
$t_{kl} =$ stress tensor,	$f_l =$ body force density,
$m_{kl} =$ couple stress tensor,	$l_l =$ body couple density,
$\epsilon =$ internal energy density,	$q_k =$ heat vector,
$\eta =$ entropy density,	$\theta =$ absolute temperature,
$h =$ the energy source,	$b =$ entropy source.

Equations (3.1)–(3.5) are, respectively, the local balance laws of mass, microinertia, momentum, moment of momentum, and energy, and (3.6) is the expression of the second law of thermodynamics. Spin inertia $\dot{\sigma}_k$ is defined by

$$\dot{\sigma}_k = \frac{D}{Dt}(j_{kl}\nu_l) = j_{kl}\dot{\nu}_l - \epsilon_{kmr}j_{lm}\nu_r\nu_l. \tag{3.7}$$

For future use we also need the integral of (3.2) as given by Eringen[10]:

$$j_{kl} = J_{KL}\chi_{kK}\chi_{lL}, \tag{3.8}$$

where J_{KL} is the microinertia tensor at the natural state of the body. For liquid crystals it may be interpreted as the moment of inertia tensor per unit mass of a molecular element (or an aggregate).

Mechanical balance laws must be supplemented by the electromagnetic balance laws.

Gauss' law:

$$\nabla\cdot\mathbf{D} = q_f, \quad \text{in } \mathscr{V} - \sigma, \tag{3.9}$$

Faraday's law:

$$\nabla\times\mathscr{E} + \frac{1}{c}\overset{*}{\mathbf{B}} = \mathbf{0}, \quad \text{in } \mathscr{V} - \sigma, \tag{3.10}$$

Magnetic flux:

$$\nabla\cdot\mathbf{B} = 0, \quad \text{in } \mathscr{V} - \sigma, \tag{3.11}$$

Ampere's law:

$$\nabla\times\mathscr{H} - \frac{1}{c}\overset{*}{\mathbf{D}} - \frac{1}{c}\mathscr{J} = \mathbf{0}, \quad \text{in } \mathscr{V} - \sigma. \tag{3.12}$$

These laws are written in a coordinate system co-moving with the points of the body so that they are Lorentz invariant to within v^2/c^2 in terms of the fields defined by

$$\mathscr{E} = \mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}, \quad \mathscr{H} = \mathbf{H} - \frac{1}{c}\mathbf{v}\times\mathbf{D},$$
$$\mathscr{J} = \mathbf{J} - q_f\mathbf{v}, \quad \mathscr{M} = \mathbf{M} + \frac{1}{c}\mathbf{v}\times\mathbf{P}, \tag{3.13}$$

where c is the velocity of light in vacuum, $\mathscr{E}$ is the electric field, $\mathscr{H}$ is the magnetic field, $\mathscr{J}$ is the conduction current, and $\mathscr{M}$ is the magnetization expressed in terms of the laboratory fields.

$\mathbf{D} =$ dielectric displacement vector,
$\mathbf{E} =$ electric field vector,

$\mathbf{B}$ = magnetic flux vector,
$\mathbf{H}$ = magnetic field vector,
$\mathbf{P}$ = polarization vector,
$\mathbf{M}$ = magnetization vector,
$\mathbf{J}$ = total current vector,
q_f = free charge density.

As we know $\mathbf{P}$ and $\mathbf{M}$ defined by

$$\mathbf{P} = \mathbf{D} - \mathbf{E}, \quad \mathbf{M} = \mathbf{B} - \mathbf{H} \tag{3.14}$$

arise from electric and magnetic multipoles. In (3.10) and (3.12) "star-derivative" occurs. This is defined by

$$\overset{*}{A}_k = \dot{A}_k + A_k v_{l,l} - A_l v_{k,l}. \tag{3.15}$$

The body force $\mathbf{f}$, body couple $\mathbf{l}$ and the energy source h is the sum of purely mechanical parts and $E - M$ parts, i.e.,

$$\rho\mathbf{f} = \rho\mathbf{f}_0 + {}_M\mathbf{f}, \quad \rho\mathbf{l} = \rho\mathbf{l}_0 + {}_M\mathbf{l}, \qquad \rho h = \rho h_0 + {}_M h, \tag{3.16}$$

where $\mathbf{f}_0$, $\mathbf{l}_0$ and h_0 are of purely mechanical origin and ${}_M\mathbf{f}$, ${}_M\mathbf{l}$ and ${}_M h$ are the contributions arising from the interaction of E-M field with the body. These are given by (cf. Maugin and Eringen[6])

$$\begin{aligned} {}_M\mathbf{f} &= q_f\mathscr{E} + \frac{1}{c}\left[\mathscr{J} + \overset{*}{\mathbf{P}}\right]\times\mathbf{B} + (\mathbf{P}\cdot\nabla)\mathscr{E} + (\nabla B)\mathscr{M}, \\ {}_M l &= \mathbf{P}\times\mathscr{E} + \mathscr{M}\times\mathbf{B}, \\ {}_M h &= \rho\mathscr{E}\cdot(\mathbf{P}/\rho)^{\cdot} - \mathscr{M}\cdot\dot{\mathbf{B}} + \mathscr{J}\cdot\mathscr{E} - {}_M\mathbf{l}\cdot\mathbf{v}. \end{aligned} \tag{3.17}$$

For some purpose we also need electromagnetic momentum $\mathbf{G}$ [which in the co-moving frame is given by $\mathscr{G} = (1/c)\mathscr{E}\times\mathscr{B}$ where $\mathscr{B} = \mathscr{H} + \mathscr{M}$] and the Poynting vector $\mathscr{S}$:

$$\mathbf{G} = \frac{1}{c}\mathbf{E}\times\mathbf{B}, \quad \mathscr{S} = c\mathscr{E}\times\mathscr{H}. \tag{3.18}$$

If we now introduce Helmholtz' free energy

$$\psi = \epsilon - \theta\eta - \frac{1}{\rho}\mathbf{P}\cdot\mathscr{E} \tag{3.19}$$

to replace ϵ in (3.5) and write the classical expression $b = h_0/\theta$ in the entropy inequality (3.6) we can eliminate h_0 between (3.5) and (3.6) to obtain the generalized *Clausius–Duhem* (C–D) *inequality* for the micropolar electromagnetic bodies:

$$\begin{aligned} &-\rho(\dot{\psi} + \eta\dot{\theta}) + t_{kl}a_{kl} + m_{kl}b_{lk} + \theta^{-1}\mathbf{q}\cdot\nabla\theta - \mathbf{P}\cdot\dot{\mathscr{E}} \\ &\quad - \mathscr{M}\cdot\dot{\mathbf{B}} - {}_M\mathbf{l}\cdot\mathbf{v} + \mathscr{J}\cdot\mathscr{E} \geqslant 0. \end{aligned} \tag{3.20}$$

This inequality places restrictions on the state of the body and it is fundamental in the development of the constitutive theory.

Jump Conditions: If the body is swept by a discontinuity surface $\sigma(t)$ moving with a velocity $\mathbf{u}$ in the direction of its positive unit normal $\mathbf{n}$ then the following jump conditions must be satisfied at the discontinuity surface $\sigma(t)$

$$[\rho(\mathbf{v} - \mathbf{u})]\cdot\mathbf{n} = 0, \tag{3.21}$$

$$[\rho j_{kl}(v_r - u_r)]n_r = 0, \tag{3.22}$$

$$[\rho(v_k - u_k)v_l - t_{kl} - ({}_M t_{kl} + u_k G_l)]\,n_k = 0, \tag{3.23}$$

$$[\rho\sigma_l(v_k - u_k) - m_{kl}]\,n_k = 0, \tag{3.24}$$

$$\begin{aligned} &[(\tfrac{1}{2}\rho\mathbf{v}\cdot\mathbf{v} + \tfrac{1}{2}\rho\sigma\cdot\mathbf{v} + \rho\epsilon + \tfrac{1}{2}\mathbf{E}\cdot\mathbf{E} + \tfrac{1}{2}\mathbf{B}\cdot\mathbf{B})(v_k - u_k) \\ &\quad - (t_{kl} + {}_M t_{kl} + u_k G_l)v_l - (q_k - \mathscr{S}_k)]\,n_k = 0, \end{aligned} \tag{3.25}$$

$$[\rho\eta(v_k - u_k) + \theta^{-1}q_k]n_k \geqslant 0. \tag{3.26}$$

These are the balance laws accompanying each one of the mechanical laws (3.1)–(3.6). In these equations we introduced Maxwell stress tensor ${}_M\mathbf{t}$, defined by

$$\begin{aligned} {}_M t_{kl} &= P_k\mathscr{E}_l - B_k\mathscr{M}_l + E_k E_l + B_k B_l \\ &\quad - \tfrac{1}{2}(\mathbf{E}\cdot\mathbf{E} + \mathbf{B}\cdot\mathbf{B} - 2\mathscr{M}\cdot\mathbf{B})\delta_{kl}. \end{aligned} \tag{3.27}$$

As discussed by Maugin and Eringen ${}_M\mathbf{t}$, ${}_M\mathbf{f}$, and $\mathbf{G}$ satisfy the identities

$${}_M t_{kl,k} - \frac{\partial G_l}{\partial t} = {}_M f_l, \quad \epsilon_{klm}\,{}_M t_{lm} = {}_M l_k. \tag{3.28}$$

For electromagnetic fields the jump conditions, on $\sigma(t)$, are

$$[\mathbf{D}]\cdot\mathbf{n} = w_f, \tag{3.29}$$

$$\mathbf{n}\times[\mathscr{E} + 1/c)\mathbf{B}\times(\mathbf{v} - \mathbf{u})] = \mathbf{0}, \tag{3.30}$$

$$[\mathbf{B}]\cdot\mathbf{n} = 0, \tag{3.31}$$

$$\mathbf{n}\times[\mathscr{H} - (1/c)\mathbf{D}\times(\mathbf{v} - \mathbf{u})] = \frac{1}{c}\mathscr{K}, \quad \mathscr{K}\cdot\mathbf{n} = 0 \tag{3.32}$$

where w_f is the surface charge density and $\mathscr{K}$ is the surface current density which is tangential to the surface σ. We note that by making σ coincide with the exterior surface of the body from the jump conditions (3.21)–(3.26) and (3.29)–(3.32) we can derive the boundary conditions.

4. CONSTITUTIVE EQUATIONS

The state of liquid crystals subject to electromechanical effects is determined by the characterization of the dependent constitutive variables (response functions)

$$\mathscr{Z} \equiv \{\psi;\eta,t_{kl},m_{kl},\mathbf{q},\mathbf{P},\mathscr{M},\mathscr{J}\} \tag{4.1}$$

as functions of certain independent variables that characterize the constitution of the body in motion. For the first-order rate-dependent, orientable, fluent materials, independent variables may be established according to axioms of causality and objectivity as (cf., Eringen[11])

$$\mathscr{Y} \equiv \{\rho^{-1},\theta,j_{kl},\gamma_{kl},a_{kl},b_{kl},\theta_{,k},\mathscr{E},\mathbf{B}\}. \tag{4.2}$$

Thus for the constitutive equations of liquid crystals we write, symbolically,

$$\mathscr{Z} = \mathscr{F}(\mathscr{Y}). \tag{4.3}$$

This means that *every* member of (4.1) is a function of *all* members of (4.2). These functions are scalar-valued functions for ψ and η; vector-valued functions for $\mathbf{q}$, $\mathbf{P}$, $\mathscr{M}$, and $\mathscr{J}$ and tensored-valued functions for t_{kl} and m_{kl}. The constitutive equations are restricted by

(i) the axiom of objectivity (material frame-indifference),
(ii) the second law of thermodynamics,
(iii) the axiom of time reversal.

The axiom of objectivity requires that the response functions remain form invariant under time-dependent rigid motions of the spatial frame of reference, with density fixed, as de-

scribed by

$$\bar{\mathbf{x}}(\mathbf{X},t)=\mathbf{Q}(t)\mathbf{x}(\mathbf{X},t)+\mathbf{c}_0(t),$$
$$\bar{\chi}_K(\mathbf{X},t)=\mathbf{Q}(t)\chi_K(\mathbf{X},t), \tag{4.4}$$

where $\mathbf{c}_0(t)$ is an arbitrary time-dependent translation and $\{\mathbf{Q}(t)\}$ represent the full group of orthogonal transformations, i.e.,

$$\mathbf{QQ}^T=\mathbf{Q}^T\mathbf{Q}=\mathbf{I},\quad \det\mathbf{Q}=\pm 1. \tag{4.5}$$

The axiom of objectivity requires that response functions transform according to

$$\bar{\mathscr{F}}(\bar{\mathscr{Y}})=\bar{\mathscr{F}}(\mathscr{Y}), \tag{4.6}$$

where $\bar{\mathscr{Y}}$ and $\bar{\mathscr{F}}$ are given by

$$\bar{\mathscr{Y}}=\{\rho^{-1},\theta,\mathbf{QjQ}^T,\mathbf{Q}\gamma\mathbf{Q}^T\det Q,\times\mathbf{QaQ}^T,\mathbf{QbQ}^T\det\mathbf{Q},\mathbf{Q}\nabla\theta,\mathbf{Q}\mathscr{E},\mathbf{QB}\det\mathbf{Q}\},$$
$$\bar{\mathscr{F}}=\{\psi,\eta,\mathbf{QtQ}^T,\mathbf{QmQ}^T\det\mathbf{Q},\mathbf{Qq},\mathbf{QP},\mathbf{Q}\mathscr{M}\det\mathbf{Q},\mathbf{Q}\mathscr{J}\}. \tag{4.7}$$

We first investigate the consequence of C–D inequality. To this end we note

$$D(\rho^{-1})/Dt=-\rho^{-2}\dot{\rho}=\rho^{-1}v_{k,k},$$
$$\dot{\gamma}_{kl}=b_{kl}+v_{kr}\gamma_{rl}+v_{lr}\gamma_{kr}-\gamma_{kr}a_{lr}. \tag{4.8}$$

In deriving $(4.8)_1$ we used (3.1) and for $(4.8)_2$ we employed $\dot{\chi}_{kK}=v_{kl}\chi_{lK}$, $D(X_{K,k})/Dt=-X_{K,r}v_{r,k}$, and (2.6). Upon substituting $\dot{\psi}$, calculated from (4.3), with $\mathscr{L}=\psi$, in the same way as argued in our previous work,[1] the C–D inequality yields:

$$\eta=-\frac{\partial\psi}{\partial\theta},\quad \frac{\partial\psi}{\partial a_{kl}}=\frac{\partial\psi}{\partial b_{kl}}=\frac{\partial\psi}{\partial\theta_{,k}}=0. \tag{4.9}$$

Since ψ must satisfy (4.6) for arbitrary rigid motions it is simple to show that

$$\frac{\partial\psi}{\partial j_{kr}}j_{lr}+\frac{\partial\psi}{\partial j_{rk}}j_{rl}-\frac{\partial\psi}{\partial\psi_{kr}}\gamma_{lr}-\frac{\partial\psi}{\partial\gamma_{rk}}\gamma_{rl}+P_k\mathscr{E}_l=\frac{\partial\psi}{\partial j_{lr}}j_{kr}+\frac{\partial\psi}{\partial j_{rl}}j_{rk}-\frac{\partial\psi}{\partial\gamma_{lr}}\gamma_{kr}-\frac{\partial\psi}{\partial\gamma_{rl}}\gamma_{rk}+P_l\mathscr{E}_k. \tag{4.10}$$

From the C–D inequality it then follows that

$${}_Dt_{kl}a_{lk}+{}_Dm_{kl}b_{lk}+\frac{1}{\theta}q_k\theta_{,k}-\left(\rho\frac{\partial\psi}{\partial\mathscr{E}}+\mathbf{P}\right)\cdot\dot{\mathscr{E}}-\rho\left(\frac{\partial\psi}{\partial\mathbf{B}}+\mathscr{M}\right)\cdot\dot{\mathbf{B}}+\mathscr{J}\cdot\mathscr{E}\geqslant 0, \tag{4.11}$$

where

$${}_Dt_{kl}=t_{lk}-{}_Et_{lk},\quad {}_Dm_{kl}=m_{kl}-{}_Em_{kl} \tag{4.12}$$

and

$${}_Et_{kl}=-\pi\delta_{kl}-{}_Em_{kr}\gamma_{rl},\quad {}_Em_{kl}=\rho\frac{\partial\psi}{\partial\gamma_{lk}},$$
$$\pi=-\frac{\partial\psi}{\partial\rho^{-1}},\quad \psi=\psi(\rho^{-1},\theta,\mathbf{j},\gamma,\mathscr{E},\mathbf{P}). \tag{4.13}$$

Here π is the thermodynamic pressure, ${}_E\mathbf{t}$ and ${}_E\mathbf{m}$ are the equilibrium parts of the stress and couple stress tensors, respectively.

Further the inequality is linear in the $\dot{\mathscr{E}}$ and $\dot{\mathbf{B}}$. If it is to be maintained for all possible values of $\dot{\mathscr{E}}$ and $\dot{\mathbf{B}}$ we must have

$$\mathbf{P}=-\rho\frac{\partial\psi}{\partial\mathscr{E}},\quad \mathscr{M}=-\rho\frac{\partial\psi}{\partial\mathbf{B}} \tag{4.14}$$

and

$${}_Dt_{kl}a_{lk}+{}_Dm_{kl}b_{lk}+(1/\theta)q_k\theta_{,k}+\mathscr{J}_k\mathscr{E}_k\geqslant 0. \tag{4.15}$$

If we assume that ${}_D\mathbf{t}$, ${}_D\mathbf{m}$, $\mathbf{q}$ and $\mathscr{J}$ are continuous functions of $\mathbf{a}$, $\mathbf{b}$, $\nabla\theta$ and $\mathscr{E}$ from (4.15) it follows that

$${}_D\mathbf{t}={}_D\mathbf{m}=\mathbf{0},\quad \mathbf{q}=\mathscr{J}=\mathbf{0}\quad\text{when}\quad \mathbf{a}=\mathbf{b}=\mathbf{0},\quad \nabla\theta=\mathscr{E}=\mathbf{0}. \tag{4.16}$$

We have therefore proved:

Theorem: Constitutive equations of liquid crystals are thermodynamically admissible if and only if they are of the forms (4.9), (4.12), (4.13), (4.14) *subject to* (4.15) [*consequently* (4.16)] *and the invariance under time reversal.*

The invariance under the axiom of time reversal imposes conditions that the free energy and the entropy production (4.15) shall be invariant when the sign of time is reversed. This is carried out in the next section.

5. POLYNOMIAL CONSTITUTIVE EQUATIONS

A. Equilibrium constitutive equations

The axiom of objectivity (4.6) places restrictions on the constitutive equations for ψ, ${}_D\mathbf{t}$, ${}_D\mathbf{m}$, $\mathbf{q}$, $\mathbf{P}$, $\mathscr{M}$ and $\mathscr{J}$. For the scalar function ψ of two second-order tensors $\mathbf{j}$, γ and two vectors $\mathscr{E}$ and $\mathbf{B}$, (4.6) implies that ψ will be a function of certain minimal number of invariants of these tensors and vectors. The complete set of invariants of two symmetric tensors $\mathbf{j}$, γ_S, one skew-symmetric tensor γ_A and two vectors

$$\mathbf{j},\ \gamma_S=\tfrac{1}{2}(\gamma+\gamma^T),\quad \gamma_A=\tfrac{1}{2}(\gamma-\gamma^T),\ \mathscr{E},\ \mathbf{B} \tag{5.1}$$

can be constructed by use of tables given by Wang[12] or Smith[13]. (Actually these lists need be augmented since γ is a relative tensor.) Because of the number of matrices and vectors involved, the list of these invariants is too long. Fortunately, in most physical problems the effects of the high degree invariants leading to nonlinar constitutive equations are negligible. We therefore retain only the invariants, each having a total degree less than or equal to three. Moreover, ψ must be invariant under the axiom of time reversal. Since $\mathbf{B}$ alters its sign with the time reversal several invariants containing the first power of $\mathbf{B}$ are excluded. In addition we assume that the natural state ($\gamma=\mathbf{0}$, $\mathscr{E}=\mathbf{0}$, $\mathbf{B}=\mathbf{0}$) is stress free. With these considerations after much effort we find that only the following list of 15 invariants contributes to the free energy:

$$I_1=\mathrm{tr}\gamma,\quad I_2=\mathrm{tr}\gamma^2,\quad I_3=\mathrm{tr}(\gamma\gamma^T),\quad I_4=\mathrm{tr}\mathbf{j},$$
$$I_5=\mathrm{tr}(\gamma\mathbf{j}),\quad I_6=\mathrm{tr}(\gamma^2\mathbf{j}),\quad I_7=\mathrm{tr}(\gamma^T\gamma\mathbf{j}),$$
$$I_8=\mathrm{tr}(\gamma\gamma^T\mathbf{j}),\quad I_9=\mathscr{E}\cdot\mathscr{E},\quad I_{10}=\mathscr{E}\cdot\mathbf{j}\mathscr{E},\quad I_{11}=\mathbf{B}\cdot\mathbf{B},$$
$$I_{12}=\mathbf{B}\cdot\mathbf{jB},\quad I_{13}=\mathrm{tr}(\gamma\mathscr{E}_D),\quad I_{14}=\mathrm{tr}(\mathbf{j}\gamma\mathscr{E}_D),$$
$$I_{15}=\mathrm{tr}(\mathbf{j}\gamma^T\mathscr{E}_D), \tag{5.2}$$

where $\mathscr{E}_{Dkl} = \epsilon_{klm}\mathscr{E}_m$. The polynomial form of ψ, up to and including all third degree terms, is therefore given by

$$\psi = \tfrac{1}{2}A_1 I_1^2 + \tfrac{1}{2}A_2 I_1^2 I_4 + A_3 I_1 I_5 + \tfrac{1}{2}A_4 I_2 + \tfrac{1}{2}A_5 I_2 I_4 + \tfrac{1}{2}A_6 I_3 + \tfrac{1}{2}A_7 I_3 I_4 + A_8 I_6 + \tfrac{1}{2}A_9 I_7 + \tfrac{1}{2}A_{10} I_8 - (1/2\rho)[(e_1 + e_2 I_4)I_9 + e_3 I_{10} + (\mu_1 + \mu_2 I_4) I_{11} + \mu_3 I_{12} + 2(e_4 + e_5 I_4)I_{13} + 2e_6 I_{14} + 2e_7 I_{15}]. \quad (5.3)$$

Using (4.13) and (4.14) we obtain

$$_E\mathbf{m}^T/\rho = [A_1 \operatorname{tr}\gamma + A_2 \operatorname{trj} \operatorname{tr}\gamma + A_3 \operatorname{tr}(\gamma\mathbf{j})]\mathbf{I} + A_3\mathbf{j}\operatorname{tr}\gamma + (A_4 + A_5 \operatorname{trj})\gamma^T + (A_6 + A_7 \operatorname{trj})\gamma + A_8(\mathbf{j}\gamma^T + \gamma^T\mathbf{j}) + A_9\gamma\mathbf{j} + A_{10}\mathbf{j}\gamma + \frac{1}{\rho} \times (e_4 + e_5 \operatorname{trj})\mathscr{E}_D + \frac{1}{\rho}e_6\mathbf{j}\mathscr{E}_D - \frac{1}{\rho}e_7\mathscr{E}_D\mathbf{j}, \quad (5.4)$$

$$P_k = [(e_1 + e_2 \operatorname{trj})\delta_{kl} + e_3 j_{kl}]\mathscr{E}_l - (e_4 + e_5 \operatorname{trj}) \times \epsilon_{kij}\gamma_{ij} - e_6\epsilon_{kij}j_{il}\gamma_{lj} + e_7\epsilon_{kij}\gamma_{il}j_{lj}, \quad (5.5)$$

$$\mathscr{M} = [(\mu_1 + \mu_2 \operatorname{trj})\mathbf{I} + \mu_3\mathbf{j}]\mathbf{B}. \quad (5.6)$$

Several observations are in order:

(a) Constitutive Eqs. (5.4)–(5.6) include all first and second degree physical phenomena that arise from the orientational and electromagnetic effects. Here the microinertia **j** play the role of "anisotropy indicator." The anisotropy changes with the rotations of the nematic elements. Thus all terms containing **j** possess the anisotropy due to intrinsic rotations.

(b) The terms involving the coefficients e_4–e_7 represent "curvature piezoelectric" effect. Because of the factor γ in these terms of (5.5) the rotation gradients produce polarization in the absence of the electric field. This possibility was recognized by Meyer.[14] In the director theory this effect is indicated by the contributions [cf., Ref. 2, Eq. (4.44)]:

$$\psi_p = -e_{1z}(\nabla\cdot\mathbf{n})\mathbf{n}\cdot\mathbf{E} - e_{3z}(\mathbf{n}\cdot\nabla\mathbf{n})\cdot\mathbf{E}, \quad (5.7)$$

where **n** is the director. In contrast to the two terms here, in (5.3) there are four terms (with coefficients of e_4 to e_7). If we pass to the director theory by writing

$$\chi_{k3} = n_k, \quad \mathbf{n}\cdot\mathbf{n} = 1, \quad j_{kL} = I_0(\delta_{kl} - n_k n_l) \quad (5.8)$$

as shown in Ref. 1, the invariants I_{13}, I_{14}, and I_{15} reduce to

$$I_{13} = \tfrac{1}{2}(\nabla\cdot\mathbf{n})\mathbf{n}\cdot\mathscr{E} - (\mathbf{n}\cdot\nabla\mathbf{n})\cdot\mathscr{E}, \quad I_{14} = I_0 I_{13}, \quad I_{15} = -\tfrac{1}{2}I_0(\nabla\cdot\mathbf{n})\mathbf{n}\cdot\mathscr{E} \quad (5.9)$$

which show that, in this special case, the three invariants coalesce to two, thus reducing the four terms involving e_4–e_7 to two as in the director theory. For the electrostatic case, in fact, we obtain the following correspondence between these coefficients and e_{1z} and e_{3z}:

$$e_{1z} = (1/2\rho)[e_4 + I_0(2e_5 + e_6 - e_7)], \quad e_{3z} = -(1/2\rho)[e_4 + I_0(2e_5 + e_6)]. \quad (5.10)$$

Based on experiments of Haas *et al.*,[15] the experimental value of e_{3z} for MBBA was estimated by Helfrich[16] as $|e_{3z}| = 7.10^{-4}$ cgs. It is clear therefore that for nematic liquids having arbitrarily shaped molecules (e.g., wedge-shaped and crescent-shaped molecules considered by Meyer) there may be curvature piezoelectric effects more general than those that can be accounted for by the director theory.

Finally from (5.4) we observe that purely electric field can produce couple stress because of its interactions with the rotation.

(c) Magnetic field is not coupled with the rotation gradients because of the nature of the time reversal property of $\mathscr{M}$ and **B**. If we consider again the special case of threadlike molecules then (5.6), by use of $(5.8)_3$ in the magnetostatic case, can be solved for **M** in terms of **H**

$$\mathbf{M} = \chi_\perp\mathbf{H} + \chi_a(\mathbf{H}\cdot\mathbf{n})\mathbf{n}, \quad (5.11)$$

where

$$\chi_\perp = (\mu_1 + 2\mu_2 I_0 + \mu_3 I_0)/(1 - \mu_1 - 2\mu_2 I_0 - \mu_3 I_0), \quad \chi_a = -\mu_3/(1 - \mu_1 - 2\mu_2 I_0 - \mu_3 I_0)(1 - \mu_1 - 2\mu_2 I_0). \quad (5.12)$$

Equation (5.11) is identical to that known in the director theory (cf., de Gennes,[3] p. 81). Consequently all magnetic effects studied by means of the director theory are contained in the present theory. Note however that, for the nematic liquids having molecules which deviate appreciably from straight threads, j_{ij} is not expressible in the form $(5.8)_3$, and that (5.11) represent only an approximation.

[While experimentally observed magnetic susceptibilities of magnetically oriented phases indicate uniaxiality such interpretations is dependent on the external characteristic length scale. It is well known that the molecular shapes of nematic liquid crystals can differ appreciably from straight threads, for example, molecules such as 4-amino -3, 4"-dinitro-*p*-terphenyl,[17] or polymeric substances which have flexibilities in the main chain that form nematic phase.[18] Even in PAA the shape of molecules are more ellipsoidal than threadlike (ratio of major to minor axes is roughtly 2.5 to 1, cf., Ref. 19). If the molecular aggregates (cybotactic groups) are considered to be microelements instead of individual molecules than the microinertia tensor will be considerably different than that of a threadlike element. There is some X-ray evidence for the existence of such aggregates in a nematic phase.[20] Recently, liquid crystals having disklike molecules have been reported.[21,22]]

B. Nonequilibrium constitutive equations

To construct the polynomial constitutive equations for the nonequilibrium fields $_D\mathbf{t}$, $_D\mathbf{m}$, q, and $\mathscr{J}$ we must construct their generators in terms of the argument vectors and tensors γ, **j**, **a**, **b**, $\nabla\theta$, $\mathscr{E}$, and **B**. The number of generators of these vectors and tensors are so many that exact form of these constitutive equations are lengthy. Fortunately the nonlinear effects arising from the products of the argument vectors and tensors are not important in most applications except possibly in the case of intense fields and high flow rates. We therefore construct only the constitutive equations that are linear in these fields and their products with **j** since **j** is the anisotropy indicator. Moreover, we exclude γ from these equations since the viscoelastic torsional effects are not considered. These effects may be important in problems where the relaxation phenomena must be taken into account. Remembering that $_D\mathbf{t}$, **j**, **a** are absolute tensors; **q**, $\mathscr{J}$. $\nabla\theta$, $\mathscr{E}$ are absolute vectors; $_D\mathbf{m}$, **b** are relative tensors, and **B**

is an axial vector, the constitutive equations read

$$
\begin{aligned}
{}_D\mathbf{t} &= [\alpha_1\ \mathrm{tra} + \alpha_2\ \mathrm{trj}\ \mathrm{tra} + \alpha_3 \mathrm{tr}(\mathbf{ja})]\mathbf{I} + \alpha_4 \mathbf{j}\ \mathrm{tra} \\
&\quad + (\alpha_5 + \alpha_6\ \mathrm{trj})\mathbf{a} + (\alpha_7 + \alpha_8\ \mathrm{trj})\mathbf{a}^T + \alpha_9 \mathbf{ja} \\
&\quad + \alpha_{10}\mathbf{aj} + \alpha_{11}\mathbf{ja}^T + \alpha_{12}\mathbf{a}^T\mathbf{j}, \qquad (5.13)\\
{}_D\mathbf{m} &= [\beta_1\ \mathrm{trb} + \beta_2\ \mathrm{trj}\ \mathrm{trb} + \beta_3\ \mathrm{tr}(\mathbf{jb})]\mathbf{I} \\
&\quad + \beta_4 \mathbf{j}\ \mathrm{trb} + (\beta_5 + \beta_6\ \mathrm{trj})\mathbf{b} + (\beta_7 + \beta_8\ \mathrm{trj})\mathbf{b}^T \\
&\quad + \beta_9 \mathbf{jb} + \beta_{10}\mathbf{bj} + \beta_{11}\mathbf{jb}^T + \beta_{12}\mathbf{b}^T\mathbf{j} \\
&\quad + [(\beta_{13} + \beta_{14}\ \mathrm{trj})\mathbf{I} + \beta_{15}\mathbf{j}](\nabla\theta)_D + \beta_{16}(\nabla\theta)_D\mathbf{j} \\
&\quad + [(\beta_{17} + \beta_{18}\ \mathrm{trj})\mathbf{I} + \beta_{19}\mathbf{j}]\mathscr{E}_D + \beta_{20}\mathscr{E}_D\mathbf{j}, \qquad (5.14)\\
\mathbf{q}/\theta &= [(\kappa_1 + \kappa_2\ \mathrm{trj})\mathbf{I} + \kappa_3\mathbf{j}]\nabla\theta + [(\kappa_4 + \kappa_5\ \mathrm{trj})\mathbf{I} + \kappa_6\mathbf{j}]\mathscr{E} \\
&\quad + (\kappa_7 + \kappa_8\ \mathrm{trj})\mathbf{b}_D + \kappa_9(\mathbf{bj})_D + \kappa_{10}(\mathbf{jb})_D \qquad (5.15)\\
\mathscr{J} &= [(\sigma_1 + \sigma_2\ \mathrm{trj})\mathbf{I} + \sigma_3\mathbf{j}]\mathscr{E} \\
&\quad + [(\sigma_4 + \sigma_5\ \mathrm{trj})\mathbf{I} + \sigma_6\mathbf{j}]\nabla\theta + (\sigma_7 + \sigma_8\ \mathrm{trj})\mathbf{b}_D \\
&\quad + \sigma_9(\mathbf{bj})_D + \sigma_{10}(\mathbf{jb})_D. \qquad (5.16)
\end{aligned}
$$

In anticipating that C–D inequality will annihilate several effects from (5.13) we have excluded terms involving **B** and products of **j** and **B**. Another important condition on these equations is imposed by the invariance of the entropy inequality under the time reversal. In classical electrodynamics, the implications of this condition was studied by several authors. For a recent treatment, we refer to Borghesani and Morro.[23]

With the reversal of time $\mathscr{E}$, **j**, and $\nabla\theta$ do not alter their signs. If the entropy inequality (4.15) is to remain in one sign, we must have

$$
\begin{aligned}
&\kappa_7 = \beta_{13}, \quad \kappa_8 = \beta_{14}, \quad \kappa_9 = \beta_{15}, \quad \kappa_{10} = \beta_{16}, \\
&\sigma_7 = \beta_{17}, \quad \sigma_8 = \beta_{18}, \quad \sigma_9 = \beta_{19}, \quad \sigma_{10} = \beta_{20}.
\end{aligned} \qquad (5.17)
$$

The time reversal also implies the Onsager reciprocal relations which place the following restrictions of α_i and β_i:

$$
\alpha_3 = \alpha_4, \quad \alpha_9 = \alpha_{10}, \quad \beta_3 = \beta_4, \quad \beta_9 = \beta_{10}. \qquad (5.18)
$$

It is interesting to observe that Eq. (5.13) is not coupled with the *E–M* fields and it is identical to that of the purely mechanical theory.[1] We notice the possibilities that temperature gradient and electric field can produce couple stress. These are the so-called hidden effects which do not contribute to the entropy production. Conversely, the rate of rotation gradient may cause heat and electric conduction irrespective of temperature gradient. These effects appear to have not been noticed before in the literature. In addition, we notice the *Seebeck effect* (coefficients of σ_4 to σ_6) and *Peltier effect* (conefficients of κ_4 to κ_6), both of which are anisotropic in that they depend on **j**.

The laws of heat conduction and electric conduction (5.15) and (5.16) are anisotropic depending on the form of **j** at each instant. The anisotropy is in general not uniaxial as in the case (5.11) and becomes so when the nematic liquids are considered threadlike. In this regard the remarks made about (5.6) also apply here.

Below we give compact forms of constitutive equations:
Equilibrium constitutive equations:

$$
\begin{aligned}
&\psi = \tfrac{1}{2}A_{ijkl}\gamma_{ij}\gamma_{kl} + E_{ijk}\gamma_{ij}\mathscr{E}_k + \tfrac{1}{2}E_{ij}\mathscr{E}_i\mathscr{E}_j + \tfrac{1}{2}M_{ij}B_iB_j, \\
&{}_E\mathbf{t} = -\pi\mathbf{I} - {}_E\mathbf{m}\boldsymbol{\gamma}, \quad \eta = -\frac{\partial\psi}{\partial\theta}, \quad \pi = -\frac{\partial\psi}{\partial\rho^{-1}}, \\
&{}_Em_{ij} = \rho A_{jikl}\gamma_{kl} + \rho E_{jik}E_k, \\
&P_k = -\rho E_{ijk}\gamma_{ij} - \rho E_{kj}E_j, \qquad (5.19)\\
&M_k = -\rho M_{kj}B_j,
\end{aligned}
$$

where

$$
\begin{aligned}
A_{ijkl} &= (A_1 + A_2\ \mathrm{trj})\delta_{ij}\delta_{kl} + A_3(j_{ij}\delta_{kl} + j_{kl}\delta_{ij}) \\
&\quad + (A_4 + A_5\mathrm{trj})\delta_{il}\delta_{jk} + (A_6 + A_7\mathrm{trj})\delta_{ik}\delta_{jl} \\
&\quad + A_8(j_{il}\delta_{jk} + j_{jk}\delta_{il}) + A_9(j_{jl}\delta_{ik} + j_{ik}\delta_{jl}), \qquad (5.20)
\end{aligned}
$$

$$
\begin{aligned}
E_{ijl} &= \frac{1}{\rho}[(e_4 + e_5\mathrm{trj})\epsilon_{ijk} + e_6\epsilon_{jkl}j_{li} + e_7\epsilon_{ikl}j_{lj}], \\
E_{ij} &= -\frac{1}{\rho}[(e_1 + e_2\mathrm{trj})\delta_{ij} + e_3 j_{ij}], \\
M_{ij} &= -\frac{1}{\rho}[(\mu_1 + \mu_2\mathrm{trj})\delta_{ij} + \mu_3 j_{ij}].
\end{aligned}
$$

Nonequilibrium constitutive equations:

$$
\begin{aligned}
&{}_Dt_{ij} = \alpha_{jikl}a_{kl}, \\
&{}_Dm_{ij} = \beta_{jikl}b_{kl} + \beta^{\theta}_{jik}\theta_{,k} + \beta^{e}_{jik}\mathscr{E}_k, \\
&q_i/\theta = \kappa_{ij}\theta_{,j} + \kappa^{e}_{ij}\mathscr{E}_j - b^{\theta}_{jki}b_{jk}, \\
&\mathscr{J}_i = \sigma_{ij}\mathscr{E}_j + \sigma^{\theta}_{ij}\theta_{,j} - \beta^{e}_{jki}b_{jk},
\end{aligned} \qquad (5.21)
$$

where

$$
\begin{aligned}
\alpha_{ijkl} &= (\alpha_1 + \alpha_2\mathrm{trj})\delta_{ij}\delta_{kl} + \alpha_3(j_{ij}\delta_{kl} + j_{kl}\delta_{ij}) \\
&\quad + (\alpha_5 + \alpha_6\mathrm{trj})\delta_{il}\delta_{jk} + (\alpha_7 + \alpha_8\mathrm{trj})\delta_{ik}\delta_{jl} \\
&\quad + \alpha_9(j_{jk}\delta_{il} + j_{il}\delta_{jk}) + \alpha_{11}j_{jl}\delta_{ik} + \alpha_{12}j_{ik}\delta_{jl}, \\
\beta^{\theta}_{ijk} &= [(\beta_{13} + \beta_{14}\mathrm{trj})\delta_{jl} + \beta_{15}j_{jl}]\epsilon_{lik} + \beta_{16}\epsilon_{jlk}j_{li}, \\
\beta^{e}_{ijk} &= [(\beta_{17} + \beta_{18}\mathrm{trj})\delta_{jl} + \beta_{19}j_{jl}]\epsilon_{lik} + \beta_{20}\epsilon_{jlk}j_{li}, \\
\kappa_{ij} &= (\kappa_1 + \kappa_2\mathrm{trj})\delta_{ij} + \kappa_3 j_{ij}, \\
\sigma_{ij} &= (\sigma_1 + \sigma_2\mathrm{trj})\delta_{ij} + \sigma_3 j_{ij}, \\
\kappa^{e}_{ij} &= (\kappa_4 + \kappa_5\mathrm{trj})\delta_{ij} + \kappa_6 j_{ij}, \\
\sigma^{\theta}_{ij} &= (\sigma_4 + \sigma_5\mathrm{trj})\delta_{ij} + \sigma_6 j_{ij},
\end{aligned}
$$

and β_{ijkl} is identical to the expression of α_{ijkl} with α_i replaced by β_i.

6. THERMODYNAMIC RESTRICTIONS

In this section we study thermodynamic restrictions arising from the entropy inequality (4.15) and the conditions under which the free energy is nonnegative. Upon employing (5.21) in (4.15) we see that the contributions **a** and **b** to the entropy production uncoupled from those of $\mathscr{E}$ and $\nabla\theta$ so that (4.15) is not violated if and only if

$$
\alpha_{ijkl}a_{ij}a_{kl} \geqslant 0, \qquad (6.1)
$$

$$
\beta_{ijkl}b_{ij}b_{kl} \geqslant 0, \qquad (6.2)
$$

$$
\kappa_{ij}\theta_{,k}\theta_{,j} + \sigma_{ij}\mathscr{E}_i\mathscr{E}_j + (\kappa^{e}_{ji} + \sigma^{\theta}_{ij})\mathscr{E}_i\theta_{,j} \geqslant 0. \qquad (6.3)
$$

The inequality (6.1) was already investigated in our previous work.[1] The necessary and sufficient conditions for (6.2) not to be violated are given by (7.12) of Ref. 1. (We note

however, that the inequalities (7.13) of Ref. 1, derived from (7.12) with extra conditions that (7.12) remain valid for all j_{ii}, are sufficient but not necessary.

The consequence of inequality (6.2) is identical to that of (6.1) with α_{ijkl} replaced by β_{ijkl}.

The inequality (6.3) may be written in the compact from

$$K_{\alpha\beta}\lambda_\alpha\lambda_\beta \geqslant 0, \quad \alpha,\beta = 1,2,\ldots,6 \tag{6.4}$$

where $K_{\alpha\beta} = K_{\beta\alpha}$ and

$$\begin{aligned} &\lambda_i = \theta_{,i}, \quad \lambda_{i+3} = \mathscr{E}_i, \\ &K_{ij} = \kappa_{ij}, \quad K_{i+3,j+3} = \sigma_{ij}, \quad i,j = 1,2,3, \\ &K_{i+3,j} = K_{j,i+3} = \tfrac{1}{2}(\kappa^e_{ji} + \sigma^\theta_{ij}). \end{aligned} \tag{6.5}$$

It is now clear that K_{ij} is a nonnegative, 6×6 matrix. Therefore, according to a theorem of algebra, the necessary and sufficient conditions for (6.4) not to be violated for all λ_i are that all sequential subdeterminants accepting the main diagonal elements as their main diagonal elements must be nonnegative, i.e.,

$$\begin{aligned} &\det K_{\alpha\beta} \geqslant 0, \quad \text{for } \alpha,\beta = 1, \\ &\qquad\qquad \text{for } \alpha,\beta = 1,2, \\ &\qquad\qquad \text{for } \alpha,\beta = 1,2,3, \\ &\qquad\qquad \text{for } \alpha,\beta = 1,2,3,4, \\ &\qquad\qquad \text{for } \alpha,\beta = 1,2,3,4,5, \\ &\qquad\qquad \text{for } \alpha,\beta = 1.2,3,4,5,6. \end{aligned} \tag{6.6}$$

Since in (6.6), the first three subdeterminants are uncoupled from the others, we see that κ_{ij} is by itself an nonnegative matrix. By exchanging $\theta_{,i}$ and $\mathscr{E}_i$ one can deduce that σ_{ij} is also a nonnegative matrix, i.e.,

$$\kappa_{ij}\theta_{,i}\theta_{,j} \geqslant 0, \quad \sigma_{ij}\mathscr{E}_i\mathscr{E}_j \geqslant 0. \tag{6.7}$$

Referred to the principal axes of j_{ij}, κ_{ij} is simplified to

$$\kappa_{ij} = (\kappa_1 + \kappa_2 \operatorname{tr}\mathbf{j} + \kappa_3 j_{ii})\delta_{ij}, \tag{6.8}$$

where underlined indices are not summed. It is now clear that the quadratic form $(6.7)_1$ is nonnegative if and only if

$$\kappa_1 + \kappa_2 \operatorname{tr}\mathbf{j} + \kappa_3 j_{ii} \geqslant 0. \tag{6.9}$$

Since $j_{ii} \geqslant 0$, (6.9) is not violated for all j_{ii} if and only if

$$\kappa_1 \geqslant 0, \quad \kappa_2 \geqslant 0, \quad 2\kappa_2 + \kappa_3 \geqslant 0. \tag{6.10}$$

[In our previous work [1] [Eq. $(6.4)_3$], $(6.10)_3$ was given as $\kappa_3 \geqslant 0$. This is necessary but not sufficient since $\operatorname{tr}\mathbf{j} = 0$ implies $j_{ii} = 0$ but the converse is not true. If we write (6.9) for $i = 1$ in the form $\kappa_1 + \kappa_2(j_{11} + j_{22} - j_{33}) + (2\kappa_2 + \kappa_3) \times j_{11} \geqslant 0$, we obtain (6.10).]

The situation for σ_{ij} is identical to (6.8)–(6.10).

For the stable equilibrium it is required that the free energy be nonnegative for all $\boldsymbol{\gamma}$, $\mathscr{E}$, and $\mathbf{B}$. This places restrictions on A_i, e_i, and μ_i in (5.3). An inspections shows that the contribution of the magnetic flux $\mathbf{B}$ to the free energy is uncoupled from those of $\boldsymbol{\gamma}$ and $\mathscr{E}$ fields. Hence the necessary and sufficient conditions for the magnetic free energy to be non-negative for all $\mathbf{B}$ is

$$\mu_1 + \mu_2 \operatorname{tr}\mathbf{j} + \mu_3 j_{ii} \leqslant 0. \tag{6.11}$$

This in turn is not violated for all j_{ii} if and only if

$$\mu_1 \leqslant 0, \quad \mu_2 \leqslant 0, \quad 2\mu_2 + \mu_3 \leqslant 0. \tag{6.12}$$

Most liquid crystals are diamagnetic. With the applied field their long axes are oriented in the direction of the field. This is indeed observed, for example, in PAA and MBBA. In order for this to occur, we must have $\mu_3 < 0$. Under the condition (6.12) this is possible since $\mu_3 \leqslant -2\mu_2$ where $-\mu_2 \geqslant 0$. It is interesting to note that this inequality does not exclude the case $\mu_3 > 0$ which also occurs in some liquid crystals where the orientation of long axes are perpendicular to the direction of the magnetic field.

To investigate the restrictions on the remainder of the free energy we write it in the compact form

$$\tfrac{1}{2}\sum_{\alpha=1}^{4} K^\alpha_{IJ}\lambda^\alpha_I\lambda^\alpha_J \geqslant 0, \tag{6.13}$$

where

$$\begin{aligned} &\lambda^1_1 = \gamma_{11}, \quad \lambda^1_2 = \gamma_{22}, \quad \lambda^1_3 = \gamma_{33}, \\ &\lambda^2_1 = \gamma_{23}, \quad \lambda^2_2 = \gamma_{32}, \quad \lambda^2_3 = \mathscr{E}_1, \\ &\lambda^3_1 = \gamma_{31}, \quad \lambda^3_2 = \gamma_{13}, \quad \lambda^3_3 = \mathscr{E}_2, \\ &\lambda^4_1 = \gamma_{12}, \quad \lambda^4_2 = \gamma_{21}, \quad \lambda^4_3 = \mathscr{E}_3, \end{aligned} \tag{6.14}$$

and $K^\alpha_{ij} = K^\alpha_{ji}$ in the principle axes of j_{ij} are given by

$$\begin{aligned} &K^1_{ij} = A_{i\underline{jj}}, \quad i,j = 1,2,3, \\ &K^2_{11} = A_{2323}, \quad K^2_{12} = A_{2332}, \quad K^2_{22} = A_{3232}, \\ &K^2_{13} = -K^2_{23} = \tfrac{1}{2}E_{231}, \quad K^2_{33} = E_{11}, \quad K^3_{11} = A_{3131}, \\ &K^3_{12} = A_{3113}, \quad K^3_{22} = A_{1313}, \\ &K^3_{13} = -K^3_{23} = \tfrac{1}{2}E_{312}, \quad K^3_{33} = E_{22}, \quad K^4_{11} = A_{1212}, \\ &K^4_{12} = A_{1221}, \quad K^4_{22} = A_{2121}, \quad K^4_{13} = -K^4_{23} = \tfrac{1}{2}E_{123}, \\ &K^4_{33} = E_{33}. \end{aligned} \tag{6.15}$$

Referred to the principal axes of $\mathbf{j}$ we have

$$\begin{aligned} A_{ijkl} = {} & [A_1 + A_2 \operatorname{tr}\mathbf{j} + A_3(j_{ii} + j_{\underline{kk}})]\delta_{ij}\delta_{kl} \\ & + [A_4 + A_5 \operatorname{tr}\mathbf{j} + A_8(j_{ii} + j_{\underline{jj}})]\delta_{il}\delta_{jk} \\ & + [A_6 + A_7 \operatorname{tr}\mathbf{j} + A_9(j_{ii} + j_{\underline{jj}})]\delta_{ik}\delta_{jl}, \end{aligned} \tag{6.16}$$

$$E_{ijk} = \frac{1}{\rho}\epsilon_{ijk}(e_4 + e_5 \operatorname{tr}\mathbf{j} + e_6 j_{ii} - e_7 j_{\underline{jj}}),$$

$$E_{ij} = -\frac{1}{\rho}(e_1 + e_2 \operatorname{tr}\mathbf{j} + e_3 j_{ii}).$$

It is now clear that the four 3×3 matrices $\mathbf{K}^\alpha$, $\alpha = 1,2,3,4$ are uncoupled. According to a theorem of algebra the necessary and sufficient conditions for $\mathbf{K}^\alpha$ to be nonnegative are

$$K^\alpha_{ii} \geqslant 0, \quad \begin{vmatrix} K^\alpha_{11} & K^\alpha_{12} \\ K^\alpha_{21} & K^\alpha_{22} \end{vmatrix} \geqslant 0, \quad \det K^\alpha_{ij} \geqslant 0 \tag{6.17}$$

for $\alpha = 1,2,3,4$. From the first one of these for $i = 3$ and $\alpha = 2,3,4$ we deduce that $e_{ii} \geqslant 0$ or

$$e_1 + e_2 \operatorname{tr}\mathbf{j} + e_3 j_{ii} \leqslant 0 \tag{6.18}$$

which is not violated for all j_{ii} if and only if

$$e_1 \leqslant 0, \quad e_2 \leqslant 0, \quad 2e_2 + e_3 \leqslant 0. \tag{6.19}$$

Similarly, from $(6.17)_1$ for $\alpha = 1,2,3,4$ we obtain

$$
\begin{aligned}
&A_1 + A_4 + A_6 + (A_2 + A_5 + A_7)\mathrm{trj} \\
&+ 2(A_3 + A_8 + A_9)j_{ii} \geqslant 0, \\
&A_6 + A_7\,\mathrm{trj} + A_9(j_{22} + j_{33}) \geqslant 0, \\
&A_6 + A_7\,\mathrm{trj} + A_9(j_{11} + j_{22}) \geqslant 0, \\
&A_6 + A_7\,\mathrm{trj} + A_9(j_{33} + j_{11}) \geqslant 0.
\end{aligned} \tag{6.20}
$$

Similarly, by expanding $(6.17)_2$ and $(6.17)_3$, one can obtain other conditions which we do not produce here, as the matter is now reduced to a routine. (A set of conditions on A_i were obtained in Ref. 1 for purely mechanical case.)

7. PASSAGE TO DIRECTOR THEORY

Most liquid crystals are constituted by straight, threadlike molecular elements. For such nematic liquids, one may simplify the basic equations by the introduction of the director concept. Suppose that Ξ_K is the common direction of the treadlike molecules at the natural state, then we may write

$$
d_k = \xi_k = \chi_{kK}\Xi_K, \quad \Xi_K = \chi_{kK}d_k, \tag{7.1}
$$

where d_k is a vector of unit magnitude called *the director*

$$
d_k d_k = 1. \tag{7.2}
$$

The independent variables appearing in the constitutive equations and the rotational inertia $\dot{\sigma}_k$ are replaced by

$$
\begin{aligned}
&\nu_{kl} = \mathring{d}_k d_l - d_k \mathring{d}_l + w_{kl}, \quad \boldsymbol{\nu} = \mathbf{d}\times\mathring{\mathbf{d}} + \tfrac{1}{2}\nabla\times\mathbf{v}, \\
&a_{kl} = d_{kl} + \mathring{d}_k d_l - d_k \mathring{d}_l, \quad b_{kl} = \nu_{k,l}, \\
&\gamma_{kl} = \epsilon_{kmn} d_m d_{n,l}, \quad j_{kl} = I_0(\delta_{kl} - d_k d_l), \\
&\dot{\sigma}_k = \frac{D}{Dt}(j_{kl}\nu_l),
\end{aligned} \tag{7.3}
$$

where I_0 is the microinertia (per unit mass) of an element about an axis through its mass center, perpendicular to the long axis. As usual, deformation and vorticity tensors $\mathbf{d}$ and $\mathbf{w}$ are given by

$$
d_{kl} = \tfrac{1}{2}(v_{k,l} + v_{l,k}), \quad w_{kl} = \tfrac{1}{2}(v_{k,l} - v_{l,k}) \tag{7.4}
$$

and the invariant time rate indicated by a superposed (°) on vectors is defined by

$$
\mathring{d}_k = \dot{d}_k - w_{kl}d_l. \tag{7.5}
$$

Employing (7.2)–(7.5) in the balance laws and constitutive equations, we obtain the field equations of nematic liquid crystals whose molecular elements are straight threadlike. These equations are now partial differential equations for the velocity field v_k and the director field d_k.

8. FREDERICKSZ' TRANSITION REVISITED

Here we discuss the rotations of nematic liquid crystals, contained between two glass slides, under the action of constant magnetic field H perpendicular to the directions of the nematic elements. This problem is known as the Fredericksz' transition, first observed by Fredericksz and Zolina. [24] The solution can be used to determine some of the material constants. Various approaches to the problem based on variational principle exist, cf., Stephen and Straley [2] and references therein. Here we give the direct treatment by use of the field equations and constitutive theory to illustrate the present development. Note also that our nematic elements are not necessarily threadlike and the method of approach is perfectly general and applicable to more complex static and dynamic problems.

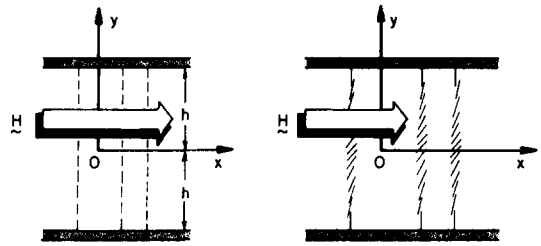

FIG. 1. Fredericks transition.

We consider the nematic elements are placed between two parallel glass plates with their long axes initially perpendicular to the glass plates. The nematic liquid is at rest ($\mathbf{v} = \mathbf{0}$) however under the influence of an applied uniform magnetic field $\mathbf{H}$, parallel to the plates, they can rotate in the (x,y) plane, Fig. 1. Thus

$$
\phi_1 = \phi_2 = 0, \quad \phi_3 = \phi(y). \tag{8.1}
$$

From (2.11) we have $n_1 = n_2 = 0, n_3 = 1$ so that (2.9) gives

$$
\gamma_{kl} = \phi'\delta_{l2}\delta_{k3}. \tag{8.2}
$$

Using (8.1) in (2.8) and (3.8) we obtain j_{kl}

$$
\begin{aligned}
j_{kl} = {} & (J^0_{kl} - J^0_{3l}\delta_{k3} - J^0_{k3}\delta_{l3} + J^0_{33}\delta_{k3}\delta_{l3})\cos^2\phi \\
& + (J^0_{3l}\delta_{k3} + J^0_{k3}\delta_{l3} \\
& - 2J^0_{33}\delta_{k3}\delta_{l3})\cos\phi + J^0_{33}\delta_{k3}\delta_{l3} \\
& + (-\epsilon_{kr3}J^0_{rl} + \epsilon_{kr3}J^0_{r3}\delta_{l3} \\
& - \epsilon_{lr3}J^0_{kr} + \epsilon_{lr3}J^0_{3r}\delta_{k3})\sin\phi\cos\phi \\
& + (-\epsilon_{kr3}\delta_{l3}J^0_{r3} - \epsilon_{lr3}\delta_{k3}J^0_{3r})\sin\phi \\
& + \epsilon_{kr3}\epsilon_{ls3}J^0_{rs}\sin^2\phi,
\end{aligned} \tag{8.3}
$$

where ϕ is polar angle measured from the x axis in counterclockwise direction and

$$
J^0_{kl} = J_{KL}\delta_{kK}\delta_{lL}. \tag{8.4}
$$

For the present configuration and rodlike elements we have $J_{KL} = 0, K \neq L$. For threadlike elements we may also approximate

$$
J^0_{11} = J^0_{33} = J^0, \quad \text{all other } J^0_{kl} = 0
$$

so that for threadlike elements (8.3) gives

$$
\begin{aligned}
&j_{11} = J_0\cos^2\phi, \quad j_{22} = J_0\sin^2\phi, \\
&j_{33} = J^0, \quad j_{12} = J_0\sin\phi\cos\phi.
\end{aligned} \tag{8.5}
$$

The couple stress and the stress tensor are given by

$$
m_{ji}/\rho = A_{ji}\phi', \quad t_{ji} = -p\delta_{ij} - \rho A_{i3}\delta_{j2}\phi'^2, \tag{8.6}
$$

where

$$
\begin{aligned}
A_{ji} = {} & A_3 j_{23}\delta_{ij} + (A_4 + A_5\,\mathrm{trj})\delta_{i2}\delta_{j3} + (A_6 + A_7\,\mathrm{trj})\delta_{i3}\delta_{j2} \\
& + A_8(j_{i2}\delta_{j3} + \delta_{i2}j_{3j}) + A_9 j_{2j}\delta_{i3} + A_{10}j_{i3}\delta_{j2}.
\end{aligned} \tag{8.7}
$$

The problem being static in nature, the dissipative stress vanishes. Substituting (8.6) into (3.3) and (3.4) with vanishing

mechanical body loads and inertia we obtain the field equations

$$p_{,i} + (\rho A_{23}\delta_{i2}\phi'^2)' - B_{j,i}M_j = 0, \tag{8.8}$$

$$(\rho A_{2i}\phi')' + \epsilon_{ij2}(-\rho A_{j3}\phi'^2) + \epsilon_{ijk}M_jB_k = 0. \tag{8.9}$$

These are field equations for the incompressible nematic liquid crystals subject to magnetic field. They are, in fact, valid for plane distortions of nematics having any orientations and any arbitrary magnetic field.

The E–M balance laws read

$$\nabla\times\mathbf{H} = \mathbf{0}, \quad \nabla\cdot\mathbf{B} = 0. \tag{8.10}$$

Constitutive equations for the $\mathbf{B}$ field are given by

$$\mathbf{M} = \mathbf{B} - \mathbf{H} = (\mu_1 + \mu_2 \mathrm{tr}\mathbf{j})\mathbf{B} + \mu_3\mathbf{j}\mathbf{B}. \tag{8.11}$$

Since $\mathbf{B}$ and $\mathbf{H}$ are assumed to depend on y only from (8.10) it follows that

$$B_2 = \text{const}, \quad H_1 = H = \text{const}, \quad H_3 = \text{const}. \tag{8.12}$$

Since $j_{13} = j_{23} = 0$, from (8.11) it follows that

$$\begin{aligned} M_1 &= B_1 - H_1 = (\mu_1 + \mu_2\mathrm{tr}\mathbf{j} + \mu_3 j_{11})B_1 + \mu_3 j_{12}B_2, \\ M_2 &= B_2 - H_2 = \mu_3 j_{21}B_1 + (\mu_1 + \mu_2\mathrm{tr}\mathbf{j} + \mu_3 j_{22})B_2, \\ M_3 &= B_3 - H_3 = (\mu_1 + \mu_2\mathrm{tr}\mathbf{j} + \mu_3 j_{33})B_3. \end{aligned} \tag{8.13}$$

The boundary conditions at $y = \pm h$ requires that

$$[H_1] = 0, \quad [H_3] = 0, \quad [B_2] = 0. \tag{8.14}$$

The first of these is satisfied since $H_1 = H = \text{const}$; the second and third can be satisfied by taking $H_3 = B_2 = 0$. This through $(8.13)_3$ gives $B_3 = M_3 = 0$ and the remaining two equations of (8.13) may be solved for

$$M_1 = \chi_1(\phi)H, \quad M_2 = -H_2 = \chi_2(\phi)H, \tag{8.15}$$

where

$$\begin{aligned} &\chi_1(\phi) = -1 + \Delta^{-1}, \quad \chi_2 = \mu_3 j_{12}/\Delta, \\ &\Delta(\phi) = 1 - \mu_1 - \mu_2\,\mathrm{tr}j - \mu_3 j_{11}. \end{aligned} \tag{8.16}$$

Equations (8.8) for $i = 1,3$ show that $p = p(y)$ and for $i = 2$ it can be integrated to give

$$p + f(\phi)\phi'^2 - \tfrac{1}{2}M_1^2 = p_0, \tag{8.17}$$

where p_0 is a constant of integration.

Equations (8.9) for $i = 1,2$ are satisfied identically and for $i = 3$ lead to

$$f(\phi)\phi'' + \tfrac{1}{2}\frac{df}{d\phi}\phi'^2 - k(\phi)H^2 = 0, \tag{8.18}$$

where we set

$$\begin{aligned} f(\phi) &= \rho A_{23} = \rho(A_6 + A_7\mathrm{tr}\mathbf{j} + A_9 j_{22} + A_{10}j_{33}) \\ &= \tfrac{1}{2}(f_0 - f_2\cos2\phi + f_3\sin2\phi), \\ A_{13} &= \rho A_9 j_{21} = \tfrac{1}{2}\frac{df}{d\phi}, \\ f_0 &= \rho[2A_6 + (2A_7 + A_9)\mathrm{tr}\mathbf{J}^0] + (2A_{10} - A_9)J^0_{33}, \\ f_2 &= \rho A_9(J^0_{11} - J^0_{22}), \quad f_3 = 2\rho A_9 J^0_{12}, \\ k(\phi) &= \mu_3 j_{12}/\Delta^2. \end{aligned} \tag{8.19}$$

The problem is now reduced to solving (8.18) for ϕ. Afterward (8.17) gives the pressure rise p. It must be noted that unlike the classical approach the pressure p cannot be ignored. Had we dropped p from (8.8) we would have arrived at a contradiction here in the sense that there would be two equations (8.15) and (8.16) for the determination of ϕ.

Differential equation (8.18) may be reduced to a first-order equation by writing

$$\tfrac{1}{2}\phi'^2 = \psi(\phi), \tag{8.20}$$

$$[f(\phi)\psi(\phi)]_{,\phi} - k(\phi)H^2 = 0. \tag{8.21}$$

By integrating once we obtain

$$2\psi = \phi'^2 = H^2F(\phi;\phi_c), \tag{8.22}$$

where ϕ_c is a critical angle at which $\phi' = 0$, and

$$F(\phi;\phi_c) = 2f^{-1}(\phi)\int_{\phi_c}^{\phi} k(\xi)d\xi. \tag{8.23}$$

Integrating (7.22) once we find the general solution, in a region of ϕ where $F > 0$

$$\pm y - y_0 = H^{-1}\int^{\phi} F^{-1/2}(\eta;\phi_c)d\eta, \tag{8.24}$$

where y_0 is a constant of integration. In the region where $F < 0$ the only solution is $\phi = \text{const}$. The two integration constants ϕ_c and y_0 are determined from the boundary conditions

$$\phi(\pm h) = \pi/2. \tag{8.25}$$

Note that the trivial solution $\phi = \pi/2$ satisfies (8.25).

The pressure is given by

$$p = p_0 + H^2[\tfrac{1}{2}\chi_1^2(\phi) - f(\phi)F(\phi;\phi_c)]. \tag{8.26}$$

From (8.3) for j_{ij} we have

$$\begin{aligned} j_{11} &= (J^0_{11} + J^0_{22}) + \tfrac{1}{2}(J^0_{11} - J^0_{22})\cos2\phi - J^0_{12}\sin2\phi, \\ j_{22} &= \tfrac{1}{2}(J^0_{11} + J^0_{22}) - \tfrac{1}{2}(J^0_{11} - J^0_{22})\cos2\phi + J^0_{12}\sin2\phi, \\ j_{33} &= J^0_{33}, \\ j_{12} &= J^0_{12}\cos2\phi + \tfrac{1}{2}(J^0_{11} - J^0_{22})\sin2\phi. \end{aligned} \tag{8.27}$$

Hence Δ and $k(\phi)$ are given by

$$\begin{aligned} \Delta(\phi) &= \Delta_{10} - \Delta_{11}\cos2\phi + \Delta_{12}\sin2\phi, \\ \Delta_{10} &= 1 - \mu_1 - \mu_2\mathrm{tr}\mathbf{J}^0 - (\mu_3/2)(J^0_{11} + J^0_{22}), \\ \Delta_{11} &= (\mu_3/2)(J^0_{11} - J^0_{22}), \quad \Delta_{12} = \mu_3 J^0_{12}, \\ k(\phi) &= \frac{\mu_3 j_{12}}{\Delta^2} = \frac{1}{2\Delta^2}\frac{d\Delta}{d\phi}. \end{aligned} \tag{8.28}$$

Consequently (8.23) gives

$$F(\phi;\phi_c) = f^{-1}(\phi)[\Delta^{-1}(\phi_c) - \Delta^{-1}(\phi)]. \tag{8.29}$$

9. THREADLIKE ELEMENTS

If the elements of nematic liquids are considered to be threadlike then we may take

$$J^0_{11} = J^0_{33} = J^0, \quad \text{all other } J^0_{kl} = 0. \tag{9.1}$$

With this $k(\phi)$ reads

$$k(\phi) = \tfrac{1}{2}\mu_3 J^0\sin2\phi\,[1 - \mu_1 - (2\mu_2 + \mu_3/2)J^0 - (\mu_3/2)J^0\cos2\phi\,]^{-2}. \tag{9.2}$$

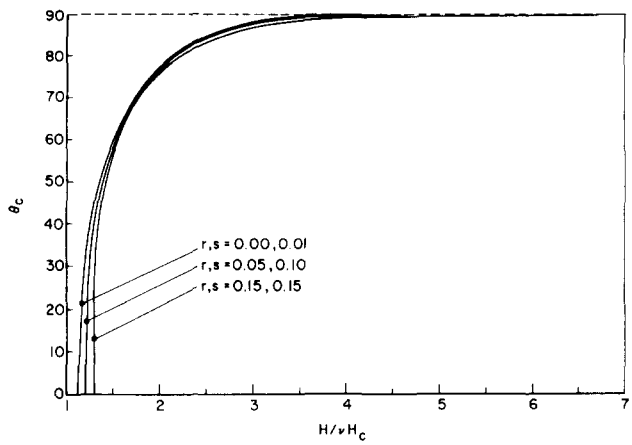

FIG. 2. Fredericks transition.

To study the angle changes from $\phi = \pi/2$ (the angles at the walls) we introduce

$$\theta = \pi/2 - \phi, \quad \theta_c = \pi/2 - \phi_c, \quad 0 \leqslant \theta \leqslant \pi. \tag{9.3}$$

First, we study the case of small angles θ which allow linearization of (8.18) so that

$$\theta'' + \lambda^2 \theta = 0, \tag{9.4}$$

where

$$\lambda^2 = 2\mu_3 J^0 H^2/(1 - \mu_1 - 2\mu_2 J^0)^2(f_0 + f_2). \tag{9.5}$$

The solution of (9.4) satisfying the boundary conditions $\theta(\pm h) = 0$ exists if

$$\sin\lambda h = 0 \quad \text{or} \quad \lambda h = n, \quad n = 0,1,2,.... \tag{9.6}$$

Excluding the trivial solution $\theta = 0$, for the critical field this gives

$$H_c = \frac{n\pi}{h}(1 - \mu_1 - 2\mu_2 J^0)\left(\frac{f_0 + f_2}{2\mu_3 J^0}\right)^{1/2} \tag{9.7}$$

Note that H_c is real since all of the quantities in the radical of (9.7) are real positive quantities. (The fact that $f(\phi) \geqslant 0$ was shown in our work [1] which expressed that $f_0 + f_2 > 0$. We also know that $\mu_3 > 0$.)

Returning to the general case (8.22) may be written as

$$2\psi = \theta'^2 = (H/\nu h H_c)/g(\theta;\theta_c,r,s) \tag{9.8}$$

where

$$\begin{aligned} &g(\theta;\theta_c,r,s) \\ &\quad = (1 + r\cos2\theta)(1 + s\cos2\theta)/(\cos2\theta - \cos2\theta_c), \\ &\nu = 2(1 + r\cos2\theta_c)/\pi^2(1 + r)^2(1 + s), \\ &r = \mu_3 J^0/(2 - 2\mu_1 - 4\mu_2 J^0 - \mu_3 J^0), \\ &s = f_2/f_0 = A_9 J^0/[2A_6 + (4A_7 + A_9 + A_{10})J^0]. \end{aligned} \tag{9.9}$$

Integrating (9.8) we obtain

$$y/h = (\nu H_c/H)G(\theta;\theta_c,r,s), \tag{9.10}$$

where

$$G(\theta;\theta_c,r,s) = \int_\theta^{\theta_c} g^{1/2}(\eta;\theta_c,r,s)d\eta, \quad 0 \leqslant y \leqslant h \tag{9.11}$$

since we expect that $\theta(-y) = \theta(y)$ and therefore $\theta'(0) = 0$ and we have $\theta_c = \theta(0)$. The remaining boundary condition at the wall gives $\theta(h) = 0$, i.e.,

$$H/\nu H_c = G(0;\theta_c,r,s) \tag{9.12}$$

and this determines θ_c. The solution in the region $-h \leqslant y < 0$ is given by $\theta(-y) = \theta(y)$.

An examination of $g(\theta;\theta_c,r,s)$ reveals that this function is nonnegative only in the region $0 \leqslant \theta \leqslant \theta_c$. For $\theta > \theta_c$ we have the only solution $\theta = \text{const}$.

It now remains to evaluate the hyperelliptic integral in (9.11). For the following two special cases (9.10) can be expressed in terms of elliptic functions.

Case (i): $s = 0$

$$\begin{aligned} y/h = &-(\nu H_c/H)\left|\tfrac{2}{r}(1 + r\cos2\theta_c)\right|^{1/2} \\ &\times[(1 - r)F(\lambda,\sin\theta_c) + r(1 + \cos2\theta_c) \\ &\times \Pi(\theta,\sin^2\theta_c, \sin\theta_c)]. \end{aligned} \tag{9.13}$$

Case (ii): $r = s$

$$\begin{aligned} y/h = &-(\nu H_c/H\sqrt{2})[(1 - r)F(\lambda, \sin\theta_c) \\ &+ r(1 + \cos2\theta_c)\cdot\Pi(\lambda,\sin^2\theta_c, \sin\theta_c), \end{aligned} \tag{9.14}$$

where F and Π are, respectively, the elliptic functions of the first and the third kind and

$$\begin{aligned} \lambda = &\arcsin(2\cos2\theta - 2\cos2\theta_c)^{1/2} \\ &\times(1 - \cos2\theta_c)^{-1/2}(1 + \cos2\theta)^{-1/2}. \end{aligned} \tag{9.15}$$

By setting $y = h$ and $\theta = 0$ in these equations we obtain θ_c as a function of $\nu H/H_c$.

We observe that classical treatment of this problem as discussed for example in Refs. 2 and 3 cannot be obtained by selecting a special value for any one of the parameters. However, if we note that $\mu_0 J_0 \ll 1$ and therefore in the expression (9.2) we may neglect $(\mu_3 J_0/2)\cos2\phi$ as compared to unity then the case (i) gives the classical result.

Finally, the case of molecules making any fixed angle other than 90° with the boundaries can be treated equally easily. In this case the boundary conditions (8.25) are replaced by $\phi(\pm h) = \phi_b$ where ϕ_b is the angle made between x axis and the long axes of the molecules at the boundary.

Calculations have been carried out to obtain the critical angle θ_c as a function of the magnetic field as given by (9.12). The results are plotted in Fig. 2, for various values of the parameters r,s. The result is as expected. In the neighborhood $H/\nu H_c \simeq 1$ magnetic elements begin to rotate severely depending on the properties, r and s, and quickly become horizontal. This result is very similar to the one already known in the director theory and observed in laboratory experiments.

ACKNOWLEDGMENT

I am indebted to Dr. F. Balta for the computer work.

[1] A.C. Eringen, "Micropolar Theory of Liquid Crystals," *Liquid Crystals and Ordered Fluids*, edited by Julian F. Johnson and Roger S. Porter (Plenum, New York, 1978), Vol. 3, pp. 443–74.

[2]M.J. Stephen and J.P. Straley, "Physics of Liquid Crystals," Rev. Mod. Phys. **46**, 617–704 (1974).
[3]P.G. deGennes, *The Physics of Liquid Crystals* (Clarendon, Oxford, 1974).
[4]R.C. Dixon and A.C. Eringen, "A Dynamical Theory of Polar Elastic Dielectric-I," Int. J. Eng. Sci. **9**, 359–77 (1965).
[5]S.R. De Grott and L.G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972).
[6]G.A. Maugin and A.C. Eringen, "On the equations of the electrodynamics of deformable bodies of finite extent," J. de Mecanique **16**, 101–47 (1977).
[7]R.A. Grot and A.C. Eringen, "Relativistic Continuum Mechanics," Int. J. Eng. Sci. **4**, 611–70 (1966).
[8]A.C. Eringen, "Linear Theory of Micropolar Elasticity," J. Math. Mech. **15**, 909–24 (1966).
[9]A.C. Eringen, *Foundations of Micropolar Thermoelasticity* (Springer-Verlag, Wien, New York, 1970).
[10]A.C. Eringen, "Simple Microfluids," Int. J. Eng. Sci. **2**, 205–17 (1964).
[11]A.C. Eringen, "A Unified Theory of Thermomechanical Materials," Int. J. Eng. Sci. **4**, 179–202 (1966).
[12]C.C. Wang, "A New Representation Theorem for Isotropic Functions: An Answer to Prof. G.F. Smith's Criticism of my paper on Representations for Isotropic Functions," Arch. Rational Mech. Anal. **36**, 166–97 (1970).
[13]G.F. Smith, "On a Fundamental Error in Two Papers of C.-C. Wang: "On Representations for Isotropic Functions, Part I and II," Arch. Rational Mech. Anal. **36**, 161–65 (1970).
[14]R.B. Meyer, "Piezoelectric Effects in Liquid Crystals," Phys. Rev. Lett. **22**, 918–21 (1969).
[15]W. Haas, J. Adams, and J.B. Flannery, "New Electro-Optic Effect in a Room-Temperature Nematic Liquid Crystals," Phys. Rev. Lett. **25**, 1326–327 (1970).
[16]W. Helfrich, "A Simple Method to Observe the Piezoelectricity of Liquid Crystals," Phys. Lett. A **35**, 393–394 (1971).
[17]P. Culling, G.W. Gray, and D. Lewis, "Mesamorphism and Polymorphism in Simple Derivatives *p*-Terphenyl," J. Chem. Soc. **2699** (1960).
[18]A. Roviello and A. Sirigu, "Mesophasic Structures in Polymers. A Preliminary Account on the Mesophases of Some Poly-Alkanoates of *p-p'*-Di-Hydroxy-α,α'-Di-Methyl Benzalazine," Polymer Lett. **13**, 455–63 (1975).
[19]I.G. Chistyakov and V.M. Chaikovskii, "Study of the Structure of Nematic *p*-Azoxyanisole in Magnetic Fields," Sov. Phys. Crystal **12**, 770–74 (1968).
[20]A. De Vries, "Evidence for the Existence of More than One Type of Nematic Phase," Mole. Crys. and Liquid Crys. **10**, 31–7 (1970).
[21]S. Chandrasekhar, B.K. Sadashva, and K.A. Surish, Pramana **9**, 471 (1977).
[22]A.M. Levlut, "Structures of a Disk-Like Mesophase," J. Physique Lett. **40**, L81 (1979).
[23]R. Borghesani and A. Morro, "Time-Reversal Invariance and Thermodynamics of Electromagnetic Fields in Materials with Memory," Annali di Matematical Pura ed applicata (VI) **IC** (1974).
[24]V. Freedericksz and V. Zolina, "Double Refraction of Thin Anisotropic Liquid Layers in the Magnetic Field," Zeits. f. Krist. **79**, 255–67 (1931).

The growth of wave discontinuities in piezoelectric semiconductors

M. F. McCarthy
National University of Ireland, University College, Galway, Ireland

H. F. Tiersten
Department of Mechanical Engineering, Aeronautical Engineering and Mechanics, Rensselaer Polytechnic Institute, Troy, New York 12181

(Received 18 July 1979; accepted for publication 22 August 1979)

The reference coordinate description of the general nonlinear differential equations describing the interaction of finitely deformable, polarizable, n-type semiconductors with the quasistatic electric field is applied in the study of acceleration waves in piezoelectric semiconductors. As a consequence, the mechanical and dielectric nonlinearities are included in the treatment as well as the semiconduction nonlinearity. The general equation for the propagation velocity of the disturbance is obtained as a function of the state of the material immediately ahead of the wavefront. In the special case of plane waves entering a homogeneous steady state, the growth equation for the amplitude of the acceleration wave is determined and, of course, the propagation velocity and coefficients in the growth equation depend on the propagation direction, but otherwise are constant. The relation between acceleration waves and the formation and propagation of acoustoelectric domains is indicated. The solutions of the growth equation indicate the formation of a shock in a finite time for conditions conducive to domain formation except in certain unusual cases possibly occurring with purely transverse acceleration waves. In the course of the treatment the condition for the threshold field for domain formation is determined under quite general circumstances. When the electrical conduction equation, which can be quite general in this treatment, is specialized to the simple form usually employed for anisotropic semiconductors, the aforementioned more general condition reduces to the anisotropic generalization of the well-known elementary result. In addition, the behavior of weak waves is discussed.

1. INTRODUCTION

In a previous investigation the theory of one-dimensional acceleration waves was applied[1] to a one-dimensional version of general rotationally invariant nonlinear electro elastic equations derived earlier from a well-defined macroscopic model[2] of deformable semiconductors. In that treatment an analytical description of the formation and propagation of purely longitudinal acoustoelectric [1]domains in piezoelectric semiconductors was obtained. The analysis indicated that for electric fields above a threshold value the amplitude of the acceleration wave would always increase without bound and become a shock. A natural and logical extension of the previous one-dimensional work is the treatment of three-dimensional acceleration waves, in which acoustoelectric domains with transverse mechanical displacement components can be considered. Recently, in the case of the quasistatic electric field the general nonlinear electroelastic equations for deformable n-type semiconductors[2] were transformed [3]from the unknown present coordinate description to the known reference coordinate description, which is the form needed here and in general for the treatment of problems.

In this paper the theory of three-dimensional acceleration waves[4–12] is applied to the above-mentioned reference coordinate description[3] of the general rotationally invariant nonlinear electroelastic equations for deformable n-type semiconductors in order to analytically describe the formation and propagation of acoustoelectric domains, with both transverse and longitudinal components of mechanical displacement, in piezoelectric semiconductors subject to high electric fields. The analysis results in an expression for the amplitude of the acceleration wave (or domain) which exhibits the competition between dissipation due to electrical conduction and the semiconduction and mechanical nonlinearities in producing decay or growth of the acceleration wave (or domain). As in the case of the purely longitudinal acceleration wave treated earlier,[1] the possibility of the amplitude of the more general three-dimensional plane acceleration wave increasing without bound and becoming a shock is clearly indicated. However, in the special case of purely transverse acceleration waves, circumstances can exist under which it is not possible for the amplitude to grow. Nevertheless, if *any* longitudinal motion is present in the acceleration wave, the possibility of the amplitude increasing without bound always exists. During the course of the analysis the expression for the velocity of the wave (or domain) as a function of the state of the material immediately ahead of the wave front naturally is obtained.

2. BASIC FORMULAS AND EQUATIONS

The macroscopic model of an elastic n-type semiconductor employed in Ref. 2 consists of three interacting, interpenetrating continua, which consist of (i) a lattice continuum which has a positive charge density; (ii) a bound electronic continuum which has a negative charge density and which can displace slightly from the lattice continuum

 0022-2488/79/122682-10$01.00

and thus produce electric polarization; and (iii) a free electronic continuum which has a negative charge density, negligible inertia, and is a conducting compressible fluid which experiences a force of resistance from its motion with respect to the lattice continuum.

Initially, the lattice continuum and the bound electronic continuum all occupy the same region of space and hence have the same reference coordinates X_L. The motion of a point of the lattice continuum is described by the mapping

$$y_i = y_i(X_L,t), \quad \mathbf{y} = \mathbf{y}(\mathbf{X},t), \tag{2.1}$$

which is one to one and differentiable as often as required. Here the y_i denote the present coordinates of material (lattice continuum) points and X_L, the reference coordinates, and t denotes the time. We consistently use the convention that capital indices denote the Cartesian components of $\mathbf{X}$ and lower case indices, the Cartesian components of $\mathbf{y}$. A comma followed by an index denotes partial differentiation with respect to a coordinate

$$g_{,i} = \frac{\partial g}{\partial y_i}(y_r,t), \quad G_{,L} = \frac{\partial G}{\partial X_L}(X_K,t), \tag{2.2}$$

and the summation convention for repeated tensor indices is employed.

Since reference coordinates are employed in our study of the propagation of acceleration waves in elastic semiconductors, the integral forms of the equations required in this work consist of Eqs. (2.41)–(2.44) of Ref. 3, which we reproduce here in the form

$$\int_{S_0} N_L(K_{Lj} + M_{Lj} - \mathscr{P}^e_{Lj})dS_0 = \frac{d}{dt}\int_{V_0} \rho_0 v_j dV_0, \tag{2.3}$$

$$\int_{S_0} N_L \mathscr{D}_L dS_0 = \int_{V_0} \bar{\mu} dV_0, \tag{2.4}$$

$$\int_{V_0} w^e_L dV_0 = \int_{S_0} N_L(\varphi + \varphi^e)dS_0, \tag{2.5}$$

$$\int_{S_0} N_L \mathscr{J}_L dS_0 = -\frac{d}{dt}\int_{V_0} \bar{\mu} dV_0, \tag{2.6}$$

where N_L denotes the outwardly directed unit normal to a reference element of area and S_0 denotes the surface enclosing the reference volume V_0. Equations (2.3)–(2.6) are the reference integral forms of the conservation of linear momentum of the combined continuum, the charge equation of electrostatics, the conservation of linear momentum of the free electronic continuum, and the conservation of total electric charge. In Eqs. (2.3)–(2.6), K_{Lj}, M_{Lj}, and $\mathscr{P}^e_{Lj}$ denote the reduced mechanical Piola–Kirchhoff stress tensor, the reference free-space Maxwell electrostatic stress tensor, and the reference free electronic pressure tensor, respectively; v_j, $\mathscr{D}_L$, w^e_L, and $\mathscr{J}_L$ denote the velocity of the solid, the reference electric displacement vector, the reference local electric field exerted on the free electronic fluid, and the reference electric current vector, respectively; ρ_0, $\bar{\mu}$, φ, and φ^e denote the reference mass density, net reference charge density, electric potential, and free electronic chemical potential, respectively; and d/dt is the material time derivative. The associated constitutive equations and additional required relations take the form

$$K_{Lj} = \rho_0 y_{j,K}\frac{\partial \chi}{\partial E_{LK}}, \quad M_{Lj} = JX_{L,i}T^{ES}_{ij},$$

$$\mathscr{P}_{Lj} = JX_{L,j}p^e, \quad \mathscr{D}_L = \epsilon_0 JX_{L,i}E_i - \rho_0\frac{\partial \chi}{\partial \mathscr{E}_L},$$

$$p^e = (\mu^e)^2\frac{\partial \epsilon^e}{\partial \mu^e}, \quad \mathscr{J}_L = J\mu^e \Omega^e_L, \tag{2.7}$$

$$J = \det y_{i,K}, \quad T^{ES}_{ij} = \epsilon_0 E_i E_j - \tfrac{1}{2}\epsilon_0 E_k E_k \delta_{ij}, \quad v_i = \frac{\partial y_i}{\partial t},$$

$$E_i = -\varphi_{,i}, \quad E_{KL} = \tfrac{1}{2}(C_{KL} - \delta_{KL}),$$
$$C_{KL} = y_{i,K}y_{i,L}, \quad \mathscr{E}_L = -\varphi_{,L}, \tag{2.8}$$

$$\chi = \chi(E_{KL},\mathscr{E}_L), \quad \epsilon^e = \epsilon^e(\mu^e), \quad \varphi^e = \partial(\mu^e\epsilon^e)/\partial\mu^e,$$
$$\Omega^e_L = \Omega^e_L(\mu^e, w^e_L, E_{LM}, \mathscr{E}_L), \tag{2.9}$$

where T^{ES}_{ij}, E_{KL}, and C_{KL} denote the free-space Maxwell electrostatic stress tensor, the material (reference or Lagrangian) strain tensor, and Green's deformation tensor, respectively; E_i, $\mathscr{E}_L$, and Ω^e_L denote the Maxwell electric field, the reference (or rotationally invariant) measure of the electric field, and the rotationally invariant constitutive vector that accounts for the relative flow velocity of the free electronic fluid, respectively; p^e, χ, μ^e, and ϵ^e denote the free electronic pressure, a particularly convenient thermodynamic state function related to the stored internal energy per unit mass of the deformable solid, the free electronic charge density, and the stored internal energy per unit charge of the free electronic fluid, respectively; and ϵ_0 is the permittivity of free space.

When the variables are appropriately differentiable, from Eqs. (2.3)–(2.6), we obtain the differential equations

$$S_{Lj,L} = \rho_0 \dot{v}_j, \tag{2.10}$$

$$\mathscr{D}_{L,L} = \bar{\mu}, \tag{2.11}$$

$$w^e_L = \varphi_{,L} + \varphi^e_{,L}, \tag{2.12}$$

$$\mathscr{J}_{L,L} + \dot{\bar{\mu}} = 0, \tag{2.13}$$

where

$$S_{Lj} = K_{Lj} + M_{Lj} - \mathscr{P}^e_{Lj}, \tag{2.14}$$

and we have employed the dot notation for partial differentiation with respect to time. We now note that we have an additional relation between the net reference charge density $\bar{\mu}$ and the free electronic charge density μ^e, which can be written in the form

$$\bar{\mu} = J\mu^e + {}_0\mu^r, \tag{2.15}$$

where ${}_0\mu^r$ is the reference residual lattice charge density, which is a constant. From Eqs. $(2.7)_6$, $(2.8)_7$, $(2.9)_{2-4}$, and (2.12), we can write

$$\mathscr{J}_L = \hat{\mathscr{J}}_L(E_{KM}, \mathscr{E}_K, G_K, \mu^e), \tag{2.16}$$

where

$$G_K = \mu^e_{,K}. \tag{2.17}$$

If we define $\hat{\chi}$ by[13]

$$\hat{\chi} = \chi - \epsilon_0 JE_k E_k/2\rho_0, \tag{2.18}$$

then by virtue of Eqs. $(2.8)_{4-7}$ and $(2.9)_1$ and the chain rule of

differentiation and the well-known relations

$$J=\sqrt{\det C_{LM}}, \quad C_{KL}^{-1}=X_{K,i}X_{L,i}, \tag{2.19}$$

we can write

$$\hat{\chi}=\hat{\chi}(E_{KL},\mathscr{E}_L). \tag{2.20}$$

Now, from Eqs. $(2.7)_{1-5}$, (2.8), $(2.9)_1$, (2.14), (2.19), and (2.20), with the aid of the differential relations

$$\frac{\partial J}{\partial(y_{i,K})}=JX_{K,i}, \quad \frac{\partial X_{N,j}}{\partial(y_{i,M})}=-X_{N,i}X_{M,j}, \tag{2.21}$$

and the chain rule of differentiation, we obtain

$$S_{Lj}=\rho_0 y_{j,K}\frac{\partial\hat{\chi}}{\partial E_{LK}}-JX_{L,j}p^e,\mathscr{D}_L=-\rho_0\frac{\partial\hat{\chi}}{\partial\mathscr{E}_L}. \tag{2.22}$$

It is clear from Eqs. $(2.7)_5$, $(2.8)_{1,5-6}$, (2.20), and (2.22) that we may write

$$S_{Lj}=\hat{S}_{Lj}(y_{j,K},\mathscr{E}_M,\mu^e),\mathscr{D}_L=\hat{\mathscr{D}}_L(E_{KM},\mathscr{E}_L), \tag{2.23}$$

and for later use we note that the constitutive response functions (2.16) and (2.23) as well as all the others are C^2 functions of their arguments.

3. GENERAL PROPERTIES OF ACCELERATION WAVES

Let σ be a propagating surface which may be represented in $(\mathbf{y},t)$ space by the equation

$$f(\mathbf{y},t)=0. \tag{3.1}$$

The unit vector $\mathbf{n}$ normal to σ and its speed of displacement u_n are given by

$$n_i=\frac{f_{,i}}{|f_{,k}|}, \quad u_n=-\frac{\partial f/\partial t}{|f_{,k}|}. \tag{3.2}$$

Corresponding to σ, we have the alternative representation of the surface Σ in $(\mathbf{X},t)$ space by means of the equation

$$\mathscr{F}(\mathbf{X},t)=f[\mathbf{y}(\mathbf{X},t),t]=0. \tag{3.3}$$

The unit vector $\mathbf{N}$ normal to Σ and its speed of propagation U_N are given by

$$N_K=\frac{\mathscr{F}_{,K}}{|\mathscr{F}_{,L}|}, \quad U_N=-\frac{\dot{\mathscr{F}}}{|\mathscr{F}_{,L}|}. \tag{3.4}$$

It is a simple matter using Eqs. (3.2)–(3.4) and the chain rule of differentiation to show that

$$N_K=\frac{F_{ik}n_i}{|F_{kL}n_k|}, \quad U_N=U\frac{|f_{,k}|}{|f_{,K}|}=U/|F_{kL}n_k|, \tag{3.5}$$

where

$$F_{iK}\equiv y_{i,K},$$

and

$$U=u_n-\mathbf{v}\cdot\mathbf{n} \tag{3.6}$$

is the local speed of propagation of the surface.

Let $\psi(\mathbf{X},t)$ be a function which suffers a jump discontinuity across the surface Σ, but is a continuous function everywhere else. We define the jump $[\psi]$ in the function ψ to be

$$[\psi]=\psi^- - \psi^+, \tag{3.7}$$

where ψ^- and ψ^+ are the limiting values of ψ immediately behind and just in front of a point lying on Σ, respectively. The surface Σ is said to be an acceleration wave if the field $y_i(\mathbf{X},t)$, $v_i(\mathbf{X},t)$, and $F_{iL}(\mathbf{X},t)$ are continuous everywhere but $\dot{v}_i(\mathbf{X},t)$, $\dot{F}_{iL}(\mathbf{X},t)$, and $F_{iL,M}(\mathbf{X},t)$, as well as all higher order partial derivatives of $y_i(\mathbf{X},t)$, suffer jump discontinuities across Σ, but are continuous functions everywhere else. From the geometric conditions of compatibility[14,15], we can obtain[16]

$$\begin{aligned}&[F_{iK,L}]=s_iN_KN_L=a_iF_{pK}F_{qL}n_pn_q, \quad s_i=[N_MN_RF_{iM,R}],\\ &[\dot{F}_{iK}]=-U_Ns_iN_K=-Ua_iF_{pK}n_p,\\ &[\dot{v}_i]=U_N^2s_i=U^2a_i, \quad s_i=B_n^2a_i, \quad B_n^2=F_{iK}F_{jK}n_in_j.\end{aligned} \tag{3.8}$$

At this point it should be noted that we need make no assumptions with regard to the continuity properties enjoyed by the electric potential $\varphi(\mathbf{X},t)=\varphi[\mathbf{y}(\mathbf{X},t),t]$ or the free electronic charge density $\mu^e(\mathbf{X},t)$ apart from assuming that at points not on Σ these functions together with their partial derivatives of all orders are continuous. The vector $\mathbf{a}$ is called the amplitude vector of the acceleration wave. If we write $\mathbf{a}=a\mathbf{r}$, where $\mathbf{r}\cdot\mathbf{n}\geqslant 0$, $|\mathbf{r}|=1$, then if $a>0$ the wave is said to be expansive, while a wave for which $a<0$ is said to be compressive. If $\mathbf{r}=\mathbf{n}$, the wave is longitudinal, while if $\mathbf{r}\cdot\mathbf{n}=0$, it is transverse.

The jump conditions across a surface of discontinuity Σ can readily be obtained from the integral forms in Eqs. (2.3)–(2.6) along with the fact that E_i remains bounded. The resulting jump conditions thus obtained consist of Eqs. $(2.46)_{1-2}$, $(2.47)_1$, (2.49) and (2.50) of Ref. 3, which are required in this work and we reproduce here in the form

$$N_k[\mathscr{J}_K]-U_N[\bar{\mu}]=0, \tag{3.9}$$

$$N_K[S_{Kj}]+\rho_0U_N[v_j]=0, \tag{3.10}$$

$$[\varphi^e]=0, \quad [\varphi]=0, \tag{3.11}$$

$$N_K[\mathscr{D}_K]=0. \tag{3.12}$$

In view of Eq. $(2.9)_3$, Eq. $(3.11)_1$ may be written

$$\hat{\varphi}^e(\mu^{e-})-\hat{\varphi}^e(\mu^{e+})=0, \tag{3.13}$$

and if we assume that $\partial\hat{\varphi}(\mu^e)/\partial\mu^e\neq 0$ it follows (the argument is given by Coleman and Gurtin[6]) that μ^e is continuous across Σ, i.e.,

$$[\mu^e]=0. \tag{3.14}$$

Furthermore, since Eq. (2.15) is of the form

$$\bar{\mu}=J\mu^e+{}_0\mu^r, \quad {}_0\mu^r=\text{constant}, \tag{3.15}$$

it follows, since $J=\det F_{iK}$ is continuous across Σ, that

$$[\bar{\mu}]=0. \tag{3.16}$$

Next, since φ is continuous across Σ, from the geometric condition of compatibility[14] and Eq. $(2.8)_7$, we have

$$[\mathscr{E}_K]=\bar{\alpha}N_K, \quad \bar{\alpha}=-[N_L\varphi_{,L}]. \tag{3.17}$$

Thus, in view of Eq. $(2.23)_2$, we may write Eq. (3.12) in the form

$$N_K\hat{\mathscr{D}}_K(E_{RL},\mathscr{E}_M^+ +\bar{\alpha}N_M)-N_K\hat{\mathscr{D}}_K(E_{RL},\mathscr{E}_M^+)=0, \tag{3.18}$$

and if we assume that $\partial\hat{\mathscr{D}}_K/\partial\mathscr{E}_L\neq 0$ it follows from Eq.

(3.18) that $\bar{\alpha}=0$ and hence from Eq. (3.17) that

$$[\mathscr{E}_K]=0, \tag{3.19}$$

and thus the reference electric field $\mathscr{E}$ is continuous across Σ. Now, in view of Eq. (3.14), from Eq. (2.17) and the geometric condition of compatibility we have

$$[G_K]=\bar{\omega}N_K, \quad \bar{\omega}=[N_L\mu^e_{,L}], \tag{3.20}$$

so that, on using Eqs. (2.16), (3.14), (3.16), (3.19), and (3.20), we may rewrite Eq. (3.9) in the form

$$N_K\hat{\mathscr{J}}_K(E_{RL},\mathscr{E}_M,G_L^+ +\bar{\omega}N_L,\mu^e) - N_K\hat{\mathscr{J}}_K(E_{RL},\mathscr{E}_M,G_L^+,\mu^e)=0. \tag{3.21}$$

From the assumption that $\partial\hat{\mathscr{J}}_K/\partial G_L\neq 0$, it follows from Eq. (3.21) that $\bar{\omega}=0$ so that

$$[G_K]=0, \tag{3.22}$$

which means that G_K is continuous across Σ. As a consequence of Eqs. (2.16), $(2.8)_{5-6}$, (2.23), (3.14), (3.19), (3.22), and the definition of an acceleration wave, we have

$$[\hat{S}_{Kj}]=[\hat{\mathscr{D}}_K]=[\hat{\mathscr{J}}_K]=0, \tag{3.23}$$

and thus Eqs. (3.9)–(3.12) are satisfied identically. Other important conditions resulting from Eqs. (3.19), (3.22), and the conditions of compatibility[14,15] are

$$[\mathscr{E}_{K,L}]=-[\varphi_{,KL}]=-\alpha N_K N_L, [\dot{\mathscr{E}}_K]=U_N\alpha N_K,$$
$$\alpha=[N_R N_S\varphi_{,RS}],$$
$$[G_{K,L}]=[\mu^e_{,KL}],=\omega N_K N_L, [\dot{G}_K]=-U_N\omega N_K,$$
$$\omega=[N_R N_S\mu^e_{,RS}]. \tag{3.24}$$

From Eqs. (2.10), (2.22), $(2.7)_{3,5}$, $(2.8)_{5-6}$, $(2.9)_{1-2}$, and (2.21), we obtain

$$A_{jKpL}F_{pL,K}+B_{jKL}\mathscr{E}_{L,K}+H_{jK}G_K=\rho_0\dot{v}_j, \tag{3.25}$$

where

$$A_{jKpL}=\frac{\partial\hat{S}_{Kj}}{\partial F_{pL}}=\rho_0\frac{\partial^2\hat{\chi}}{\partial F_{jK}\partial F_{pL}}-\frac{\partial}{\partial F_{pL}}(JX_{K,j})\hat{p}_e(\mu^e)$$
$$=\delta_{jp}S_{KL}+F_{jN}F_{pS}c_{KNLS}-J(X_{K,j}X_{L,p}-X_{L,j}X_{K,p})\times\hat{p}^e(\mu^e), \tag{3.26}$$

$$B_{jKL}=\frac{\partial\hat{S}_{Kj}}{\partial\mathscr{E}_L}=\rho_0\frac{\partial^2\hat{\chi}}{\partial F_{jK}\partial\mathscr{E}_L}=-\frac{\partial\hat{\mathscr{D}}_L}{\partial F_{jK}}=F_{jN}B_{KNL}, \tag{3.27}$$

$$H_{jK}=\frac{\partial\hat{S}_{Kj}}{\partial\mu^e}=-JX_{K,j}\frac{\partial p^e}{\partial\mu^e}(\mu^e), \tag{3.28}$$

and

$$S_{KL}=\rho_0\frac{\partial\hat{\chi}}{\partial E_{KL}}, \quad c_{KNLS}=\rho_0\frac{\partial^2\hat{\chi}}{\partial E_{KN}\partial E_{LS}},$$
$$B_{KNL}=\rho_0\frac{\partial^2\hat{\chi}}{\partial\mathscr{E}_L\partial E_{KN}}. \tag{3.29}$$

Since $\hat{S}_{Kj}(.,.,.)$ is by hypothesis a C^2 function, it follows that the coefficients in Eqs. (3.26)–(3.28) are continuous across the singular surface Σ. Thus, taking the jump in Eq. (3.25) across Σ and employing Eqs. $(3.8)_{1,4}$, (3.22), and $(3.24)_1$, we obtain

$$A_{jKpL}N_K N_L s_p-\alpha B_{jKL}N_K N_L=\rho_0 U_N^2 s_j. \tag{3.30}$$

We now need the expression for α in terms of $\mathbf{s}$. From Eqs. (2.11) and (3.27) we have

$$-B_{iLK}F_{iL,K}+\kappa_{KL}\mathscr{E}_{L,K}=\bar{\mu}, \tag{3.31}$$

where

$$\kappa_{KL}=\frac{\partial\hat{\mathscr{D}}_K}{\partial\mathscr{E}_L}=-\rho_0\frac{\partial^2\hat{\chi}}{\partial\mathscr{E}_K\partial\mathscr{E}_L}. \tag{3.32}$$

Since $\hat{\mathscr{D}}_K(.,.)$ is a C^2 function, it follows that the coefficients in Eq. (3.31) are continuous across the singular surface Σ. Thus, on taking the jump in Eq. (3.31) across Σ and using Eqs. $(3.8)_1$, (3.16), and $(3.24)_1$, we find

$$\alpha=-\zeta L_i s_i, \tag{3.33}$$

where

$$L_i=B_{iLK}N_L N_K=F_{iM}B_{MLK}N_L N_K, \quad \zeta=(\kappa_{KL}N_K N_L)^{-1}. \tag{3.34}$$

The substitution of Eq. (3.33) into (3.30) with the aid of Eqs. (3.26) and $(3.34)_1$ yields the following propagation condition:

$$(\hat{Q}_{jp}-\rho_0 U_N^2\delta_{jp})s_p=0, \tag{3.35}$$

where

$$\hat{Q}_{jp}=\hat{Q}_{jp}(F_{rL},\mathscr{E}_K,N_M)=A_{jKpL}N_K N_L+\zeta L_j L_p$$
$$=\delta_{jp}S_{KL}N_K N_L+F_{jL}F_{pM}c_{KLRM}N_K N_R+\zeta L_j L_p=\hat{Q}_{jp} \tag{3.36}$$

is the acoustic tensor. We note that $\hat{Q}_{jp}$ is symmetric and for fixed $\mathbf{n}$ it is a function of the deformation gradient F_{rL}, the reference electric field $\mathscr{E}_K$ at the wave front, but is independent of μ^e, the density of free electronic charge at the wave front. It follows from Eq. (3.35) that the amplitude $\mathbf{a}$ of an acceleration wave traveling in the direction $\mathbf{n}$ in a piezoelectric semiconductor must be a proper eigenvector of the symmetric acoustic tensor $\hat{Q}_{jp}$ and the speed of propagation U_N must be such that $\rho_0 U_N^2$ is the corresponding eigenvalue of $\hat{Q}_{jp}$.

The equation (3.35) was derived by Truesdell[17] for acceleration waves in elastic media. It has since been derived by a number of authors for acceleration waves in a variety of media.[7–9] We note in particular that the acoustic tensor (3.36) has precisely the same form as the corresponding acoustic tensor which occurs in the theory of wave propagation in elastic dielectrics.[10]

Equations (3.35) admit a nontrivial solution if and only if

$$\det(\hat{Q}_{jp}-\rho_0 U_N^2\delta_{jp})=0, \tag{3.37}$$

and this equation determines the possible speeds of propagation for a given direction of propagation $\mathbf{n}$. On the other hand, if the amplitude $\mathbf{a}$ of a wave is known, then the corresponding speed of propagation is determined by the formula

$$\rho_0 U_N^2=\hat{Q}_{jp}a_j a_p/a_i a_i. \tag{3.38}$$

Since the acoustic tensor $\hat{Q}_{jp}$ is symmetric, it has three real eigenvalues. However, at this stage it is possible that all of these eigenvalues may be negative, in which case no real waves will exist at all. We now wish to record the conditions which guarantee the possible existence of at least some real

waves. A detailed analysis of the situation for purely elastic materials has been given by Truesdell,[18] Truesdell and Noll,[19] Wang and Truesdell,[20] and Chadwick and Currie.[21]

Once the deformation gradient and electric field ahead of the wave are known, it follows from Eqs. $(3.5)_1$, (3.29), (3.34), (3.36), $(2.8)_{5-6}$, and $(2.9)_1$ that the acoustic tensor depends on $\mathbf{n}$ only, i.e., we have

$$\widehat{Q}_{jp}(F_{rL},\mathscr{E}_K,N_M)=\overline{Q}_{jp}(\mathbf{n}), \tag{3.39}$$

for fixed F_{rL} and $\mathscr{E}_K$. If

$$\overline{Q}_{ij}(\mathbf{n})n_in_j>0 \tag{3.40}$$

for all unit vectors $\mathbf{n}$, the material may be said to have positive longitudinal piezoelectricity. Truesdell[18] has shown that when Eq. (3.40) is satisfied, there exists at least one direction in which a longitudinal wave may exist and propagate. If the acoustic tensor is strongly elliptic in the sense that

$$\overline{Q}_{ij}(\mathbf{n})\mu_i\mu_j>0, \tag{3.41}$$

for all unit vectors $\mathbf{n}$ and $\boldsymbol{\mu}$, it then follows (Truesdell[18]) that there is at least one direction of propagation in which a longitudinal wave and two transverse waves with orthogonal amplitudes may exist and propagate. In particular, it should be noted that if the strong ellipticity condition (3.41) is satisfied and if the deformation and electric field are uniform ahead of the wave, then there exists at least one direction in which a plane longitudinal and two plane transverse waves may exist and propagate for all times. For propagation in all other directions under the above-mentioned circumstances, the three plane waves are, of course, not necessarily either purely longitudinal or purely transverse, but may consist of an admixture of all mechanical displacement components. Nevertheless, in the most general case, if the propagation velocities are distinct, the three plane waves have mutually orthogonal mechanical displacement fields. In the next section, we examine the manner in which the amplitudes of such waves vary as they traverse the material.

4. GROWTH AND DECAY OF PLANE ACCELERATION WAVES

In this section we derive the differential equation which determines the manner in which the amplitude of a plane acceleration wave varies as it traverses the material. It is assumed that the material ahead of the plane wave front is at rest in a state of homogeneous strain, is subject to a uniform electric field, and that the charge density of the free electronic fluid is uniform and constant prior to the arrival of the wave front.

The differentiation of the equation of motion (3.25) with respect to t, with $\mathbf{X}$ fixed, yields

$$A_{jKpL}\dot{F}_{pL,K}+B_{jKL}\dot{\mathscr{E}}_{L,K}+H_{jK}\dot{G}_K+C_j=\rho_0\ddot{v}_j, \tag{4.1}$$

where

$$\begin{aligned}C_j=&A_{jKpLqM}F_{pL,K}\dot{F}_{qM}+B_{jKpLM}F_{pL,K}\dot{\mathscr{E}}_M\\&+H_{jKpL}F_{pL,K}\dot{\mu}^e+B_{jKpML}\mathscr{E}_{L,K}\dot{F}_{pM}\\&+B_{jKLM}\mathscr{E}_{L,K}\dot{\mathscr{E}}_M+H_{jKpL}G_K\dot{F}_{pL}+\widetilde{H}_{jK}G_K\dot{\mu}^e,\end{aligned} \tag{4.2}$$

with

$$\begin{aligned}A_{jKpLqM}&=\partial^2\widehat{S}_{Kj}/\partial F_{pL}\partial F_{qM}\\&=\delta_{jp}F_{qN}c_{KLMN}+\delta_{jq}F_{pN}c_{KMLN}\\&\quad+\delta_{pq}F_{jN}c_{KNLM}\\&\quad+F_{jN}F_{pT}F_{qR}C_{KNLTMR}\\&\quad-JX_{M,q}(X_{K,j}X_{L,p}-X_{L,j}X_{K,p})p^e\\&\quad+J(X_{K,q}X_{M,j}X_{L,p}+X_{K,j}X_{M,p}X_{L,q}\\&\quad-X_{L,q}X_{M,j}X_{K,p}-X_{L,j}X_{K,q}X_{M,p})p^e,\end{aligned} \tag{4.3}$$

$$B_{jKpLM}=\frac{\partial^2\widehat{S}_{Kj}}{\partial F_{pL}\partial\mathscr{E}_M}=\delta_{jp}B_{KLM}+F_{jN}F_{pS}B_{KNLSM},$$

$$H_{jKpL}=\frac{\partial^2\widehat{S}_{Kj}}{\partial\mu^e\partial F_{pL}}=-J(X_{K,j}X_{L,p}-X_{L,j}X_{K,p})\frac{\partial\hat{p}^e}{\partial\mu^e},$$

$$B_{jKLM}=\frac{\partial^2\widehat{S}_{Kj}}{\partial\mathscr{E}_L\partial\mathscr{E}_M}=F_{jN}B_{NKLM},$$

$$\widetilde{H}_{jK}=-JX_{K,j}\frac{\partial^2\hat{p}_e}{\partial(\mu^e)^2},$$

and

$$\begin{aligned}C_{KLMNRT}&=\frac{\partial^2\widehat{S}_{KL}}{\partial E_{MN}\partial E_{RT}}=\rho_0\frac{\partial^3\hat{\chi}}{\partial E_{KL}\partial E_{MN}\partial E_{RT}},\\B_{KNLSM}&=\frac{\partial^2\widehat{S}_{KN}}{\partial E_{LS}\partial\mathscr{E}_M}=\rho_0\frac{\partial^3\hat{\chi}}{\partial E_{KN}\partial E_{LS}\partial\mathscr{E}_M},\\B_{NKLM}&=\frac{\partial^2\widehat{S}_{NK}}{\partial\mathscr{E}_L\partial\mathscr{E}_M}=\rho_0\frac{\partial^3\hat{\chi}}{\partial E_{NK}\partial\mathscr{E}_L\partial\mathscr{E}_M}.\end{aligned} \tag{4.4}$$

In view of our earlier assumptions on the continuity of the response functions $\widehat{S}_{KJ}(.,.,.)$, the coefficients in Eq. (4.3) are continuous across the singular surface Σ. On taking the jump in Eq. (4.1) across Σ, keeping in mind our assumptions concerning the uniformity of the rest state of the material ahead of the wave front and making use of the relation

$$[\varphi\psi]=\varphi^+[\psi]+\psi^+[\varphi]+[\varphi][\psi], \tag{4.5}$$

and employing Eq. (3.22) we obtain

$$\begin{aligned}&A_{jKpL}[\dot{F}_{pL,K}]+B_{jKL}[\dot{\mathscr{E}}_{L,K}]+H_{jK}[\dot{G}_K]+[C_j]\\&\quad=\rho_0[\ddot{v}_j],\end{aligned} \tag{4.6}$$

where

$$\begin{aligned}[C_j]=&A_{jKpLqM}[F_{pL,K}][\dot{F}_{qM}]+B_{jKpLM}\{[F_{pL,K}][\dot{\mathscr{E}}_M]\\&+[\dot{F}_{pL}][\mathscr{E}_{M,K}]\}+B_{jKLM}[\mathscr{E}_{L,K}][\dot{\mathscr{E}}_M].\end{aligned} \tag{4.7}$$

Since we are dealing with a plane wave front entering a uniform state, from the compatibility conditions and the definition of an acceleration wave, we have[22]

$$\begin{aligned}&[\dot{F}_{rL,K}]=b_rN_LN_K,\quad b_r=[N_MN_S\dot{F}_{rM,S}],\\&[\ddot{v}_i]=U_N^2b_i+2U_N^2\frac{\delta s_i}{\delta t},\\&[\dot{\mathscr{E}}_{L,K}]=\beta N_LN_K,\beta=-[N_RN_S\dot{\varphi}_{,RS}],\end{aligned} \tag{4.8}$$

where $\delta f/\delta t$ is the displacement derivative[23] of f. Substituting from Eqs. (3.8), (3.24), and (4.8) into Eq. (4.6), we obtain

$$(A_{jKpL}N_KN_L-\rho_0U_N^2\delta_{jp})b_p+\beta B_{jKL}N_KN_L-U_N\omega H_{jK}N_K$$

$$-2\rho_0 U_N^2 \frac{\delta s_j}{\delta t} - U_N A_{jKpLqM} N_K N_L N_M s_p s_q + 2U_N \alpha B_{jKpLM} N_K N_L N_M s_p - U_N \alpha^2 B_{jKLM} N_K N_L N_M = 0. \tag{4.9}$$

We now need the expression for ω in terms of **s**. To this end, with the aid of Eqs. (2.16) and $(2.8)_{5-6}$, we rewrite Eq. (2.13) in the form

$$\Sigma_{KrL} F_{rL,K} + \Sigma_{KL} \mathscr{E}_{L,K} + \Omega_{KL} G_{L,K} + \xi_K G_K + \dot{\tilde{\mu}} = 0, \tag{4.10}$$

where

$$\Sigma_{Krl} = \frac{\partial \hat{\mathscr{J}}_K}{\partial F_{rL}} = F_{rN} \Sigma_{KLN}, \Sigma_{KMN} = \frac{\partial \hat{\mathscr{J}}_K}{\partial E_{MN}},$$
$$\Sigma_{KL} = \frac{\partial \hat{\mathscr{J}}_K}{\partial \mathscr{E}_L}, \quad \Omega_{KL} = \frac{\partial \hat{\mathscr{J}}_K}{\partial G_L}, \xi_K = \frac{\partial \hat{\mathscr{J}}_K}{\partial \mu^e}. \tag{4.11}$$

In view of the continuity assumptions on $\hat{\mathscr{J}}_K(.,.,.,.)$, it follows that the coefficients in Eqs. (4.11) are continuous across the singular surface. Hence, on taking the jump in Eq. (4.10) across Σ and employing Eqs. (3.8), (3.22), and $(3.24)_{1,4}$, we have

$$\omega = \nu(-\Sigma_{KrL} N_L N_K s_r + \alpha \Sigma_{KL} N_K N_L - [\dot{\tilde{\mu}}]), \tag{4.12}$$

where

$$\nu = (\Omega_{KL} N_L N_K)^{-1}. \tag{4.13}$$

Taking the material time derivative of Eq. (2.15), we obtain

$$\dot{\tilde{\mu}} = J\dot{\mu}^e + \dot{J}\mu^e, \tag{4.14}$$

the jump in which, with Eqs. $(2.8)_1$, $(3.8)_3$, (3.14), (3.22), the kinematic condition of compatibility,[15] and the definition of an acceleration wave, yield

$$[\dot{\tilde{\mu}}] = -\mu^e J X_{K,i} U_N s_i N_K, \tag{4.15}$$

where we have employed the well-known relation

$$\dot{J} = J X_{K,i} \dot{F}_{iK} \tag{4.16}$$

in obtaining Eq. (4.15). The substitution of Eqs. (3.33) and (4.15) into Eq. (4.12) enables us to write

$$\omega = \nu \Sigma_r s_r, \tag{4.17}$$

where

$$\Sigma_r = -\Sigma_{KrL} N_L N_K - \xi \Sigma_{KL} N_K N_L L_r + J U_N \mu^e N_K X_{K,r}. \tag{4.18}$$

We now must express β in terms of **s** and **b**. To this end we take the material time derivative of Eq. (3.31) to obtain

$$-B_{iLK} \dot{F}_{iL,K} + \varkappa_{KL} \dot{\mathscr{E}}_{L,K} - B_{iLrSK} \dot{F}_{rS} F_{iL,K} - B_{iLKM} (\dot{\mathscr{E}}_M F_{iL,K} + \dot{F}_{iL} \mathscr{E}_{M,K}) + \varkappa_{KLM} \dot{\mathscr{E}}_M \mathscr{E}_{L,K} = \dot{\tilde{\mu}} \tag{4.19}$$

where

$$\varkappa_{KLM} = \frac{\partial^2 \hat{\mathscr{D}}_K}{\partial \mathscr{E}_L \partial \mathscr{E}_M} = -\rho_0 \frac{\partial^3 \hat{\chi}}{\partial \mathscr{E}_K \partial \mathscr{E}_L \partial \mathscr{E}_M}, \tag{4.20}$$

which is continuous across the singular surface. Clearly, all the coefficients occurring in Eq. (4.19) are continuous across the singular surface Σ. Thus, on taking the jump in Eq. (4.19) across Σ, recalling the time independent uniform state ahead of the wave front, and employing Eq. (4.15), we have

$$-B_{iLK}[\dot{F}_{iL,K}] + \varkappa_{KL}[\dot{\mathscr{E}}_{L,K}] - B_{iLrSK}[\dot{F}_{rS}][F_{iL,K}] - B_{iLKM}([\dot{\mathscr{E}}_M][F_{iL,K}] + [\dot{F}_{iL}][\mathscr{E}_{M,K}]) + \varkappa_{KLM}[\dot{\mathscr{E}}_M][\mathscr{E}_{L,K}] = [\dot{\tilde{\mu}}]. \tag{4.21}$$

Substituting from Eqs. (3.8), (3.24), (3.33), (3.34), (4.8), and (4.15) into Eq. (4.21), we obtain

$$\beta = \xi L_p b_p - \xi U_N J \mu^e X_{K,i} N_K s_i + U_N \xi M_{ij} s_i s_j, \tag{4.22}$$

where

$$M_{ij} = (-B_{iLjMK} + 2\xi B_{iLKM} L_j + \xi^2 \varkappa_{LMK} L_i L_j) \times N_K N_L N_M. \tag{4.23}$$

We now substitute from Eqs. (3.33), (4.17), and (4.22) into Eq. (4.9) and employ Eqs. $(3.34)_1$ and (3.36) to obtain

$$(\hat{Q}_{jp} - \rho_0 U_N^2 \delta_{jp}) b_p = 2\rho_0 U_N^2 \frac{\delta s_j}{\delta t} + \alpha_{jp} s_p + \alpha_{jpq} s_p s_q, \tag{4.24}$$

where

$$\alpha_{jp} = \xi U_N J \mu^e L_j X_{K,p} N_K + \nu U_N H_{jK} N_K \Sigma_p, \tag{4.25}$$

and

$$\alpha_{jpq} = U_N \xi L_j M_{pq} + U_N [A_{jKpLqM} + 2\xi B_{jKpLM} L_q + \xi^2 B_{jKLM} L_p L_q] N_K N_L N_M. \tag{4.26}$$

Note that Eqs. (4.24) serve to determine the components of the amplitude **b** of the third order discontinuity induced by the acceleration wave. However, for the moment, our primary objective is to use Eqs. (4.24) to obtain the differential equation governing the evolutionary behavior of the amplitude **a** of the acceleration wave. In Sec. 6 we study the solution of Eq. (4.24) in greater detail and we discuss both induced discontinuities and higher order waves.

Using Eqs. (3.5), $(3.8)_{4-6}$, (4.23), (4.25), and (4.26), we may rewrite Eq. (4.24) in the form

$$(\hat{Q}_{jp} - \rho_0 U_N^2 \delta_{jp}) b_p = 2\rho_0 U^2 \frac{\delta a_j}{\delta t} + \bar{\alpha}_{jp} a_p + \bar{\alpha}_{jpq} a_p a_q, \tag{4.27}$$

where

$$\bar{\alpha}_{jp} = \xi U J \mu^e L_j n_p + \nu U H_{jK} F_{sK} n_s \Sigma_p, \tag{4.28}$$
$$\bar{\alpha}_{jpq} = U \xi L_j \bar{M}_{pq} + U [A_{jKpLqM} + 2B_{jKpLM} L_q + \xi^2 B_{jKLM} L_p L_q] F_{rK} F_{sL} F_{tM} n_r n_s n_t, \tag{4.29}$$

and

$$\bar{M}_{pq} = [-B_{pKqLM} + 2\xi B_{pKLM} L_q + \xi^2 \varkappa_{KLM}] \times F_{rK} F_{sL} F_{tM} n_r n_s n_t. \tag{4.30}$$

We may write

$$a_j = a r_j, \quad |\mathbf{r}| = 1, \tag{4.31}$$

where **r** is the unit eigenvector of $\hat{Q}_{jp}$ corresponding to the eigenvalue $\rho_0 U_N^2$, and recall that, since we are dealing with a plane wave front propagating into a uniform region, the components of **r** are constant. Equation (4.27) may now be written

$$(\hat{Q}_{jp} - \rho_0 U_N^2 \delta_{jp}) b_p = 2\rho_0 U^2 r_j \frac{\delta a}{\delta t} + (\bar{\alpha}_{jp} r_p) a + (\bar{\alpha}_{jpq} r_p r_q) a^2. \tag{4.32}$$

If we now contract Eq. (4.32) with r_j and employ Eqs. (3.35) and (3.36), we find that the amplitude a of the acceleration wave satisfies the equation

$$\frac{\delta a}{\delta t} = -\omega_0 a + \beta_0 a^2, \tag{4.33}$$

where

$$\omega_0 = \frac{\bar{\alpha}_{jp} r_j r_p}{2\rho_0 U^2}, \tag{4.34}$$

$$\beta_0 = \frac{-\bar{\alpha}_{jpq} r_j r_p r_q}{2\rho_0 U^2}. \tag{4.35}$$

The implications of Eq. (4.33) are examined in detail in the next section.

5. IMPLICATIONS OF THE GROWTH EQUATION

Equation (4.33), which governs the evolutionary behavior of plane acceleration waves oriented arbitrarily and propagating into a region subject to a time independent state of homogeneous strain and uniform electric field in a deformable semiconductor, is of the same form as Eq. (5.6) of Ref. 1. Clearly, we expect the behavior of the amplitude of a plane acceleration wave to be, at least qualitatively, similar to that of a purely longitudinal wave. This is for the most part the case but, as expected, in contrast to the situation prevailing in the one-dimensional case discussed in Ref. 1, the coefficients ω_0 and β_0, defined by Eqs. (4.34) and (4.35), respectively, are not absolute constants for a given material and state even though $\mathscr{E}$, $\mathbf{F}$, and μ^e are uniform ahead of the wave front, but vary with the propagation direction $\mathbf{n}$. Of course, once $\mathscr{E}$, $\mathbf{F}$, and μ^e are prescribed ahead of the wave, then for a given $\mathbf{n}$, $\mathbf{r}$ is determined by $\mathbf{n}$ through Eqs. (3.35). After the unit vector $\mathbf{r}$ has been determined from Eqs. (3.35), the coefficients ω_0 and β_0 are fixed. Thus, for a given state ahead of the wavefront, i.e., values of $\mathbf{F}$, $\mathscr{E}$, and μ^e, ω_0 and β_0 are constants for a given $\mathbf{n}$.

When neither of the quantities ω_0 and β_0 vanishes, then the solution of Eq. (4.33) is

$$a(t) = \lambda_0 / [((\lambda_0/a_0) - 1)e^{\omega_0 t} + 1], \tag{5.1}$$

where

$$\lambda_0 = \omega_0/\beta_0, \tag{5.2}$$

and $a(0)$ is the value of the amplitude of the wave at time $t = 0$. It is clear that the behavior of the amplitude of a given plane wave is determined by the coefficients ω_0 and β_0 as well as by the initial amplitude a_0. In order to discuss all possible cases which may arise we first suppose that

$$\omega_0 > 0, \quad \beta_0 \neq 0; \tag{5.3}$$

then, in view of Eq. (5.2), we have

$$\operatorname{sgn}\lambda_0 = \operatorname{sgn}\beta_0, \tag{5.4}$$

and then from Eq. (5.1) three distinct possibilities arise:

(i) If $\operatorname{sgn}a(0) = \pm \operatorname{sgn}\beta_0$ and $|a_0| < |\lambda_0|$, then $a(t) \to 0$ monotonically as $t \to \infty$.

(ii) If $a(0) = \lambda_0$, then $a(t) = a(0)$.

(iii) If $\operatorname{sgn}a(0) = \operatorname{sgn}\beta_0$ and $|a(0)| > |\lambda_0|$, then $a(t) \to \infty$ monotonically within a finite time t_∞ given by

$$t_\infty = -(1/\omega_0) \ln\{1 - [\lambda_0/a(0)]\}. \tag{5.5}$$

We now suppose that

$$\omega_0 < 0, \quad \beta_0 \neq 0; \tag{5.6}$$

then we have

$$\operatorname{sgn}\lambda_0 = -\operatorname{sgn}\beta_0, \tag{5.7}$$

and again from Eq. (5.1) three distinct possibilities arise:

(i)* If $\operatorname{sgn}a(0) = -\operatorname{sgn}\beta_0$ and $|\lambda_0| > |a(0)|$, then $a(t) \to \lambda_0$ monotonically as $t \to \infty$.

(ii)* If $a(0) = \lambda_0$, then $a(t) = a(0)$.

(iii)* If $\operatorname{sgn}a(0) = \operatorname{sgn}\beta_0$, then $a(t) \to \infty$ within a finite time t_∞ given by

$$t_\infty = -(1/\omega_0) \ln(1 + |\lambda_0/a_0|). \tag{5.8}$$

It is clear from the foregoing results that the number λ_0 plays a fundamental role in determining whether the amplitude of an acceleration wave will grow or decay as the wave transverses the material. For this reason we follow the usual custom and call λ_0 the critical amplitude for acceleration waves encountering a homogeneous steady state. We note that if $\omega_0 > 0$, then the behavior of an acceleration wave propagating into a piezoelectric semiconductor which is in a uniform steady state is precisely the same as that of an acceleration wave propagating into a homogeneously deformed material with memory.[6,7] In particular, if the initial amplitude of the wave is less in absolute value than the critical initial amplitude, the amplitude decreases to zero as the wave propagates. On the other hand, if the initial amplitude is greater in absolute value than the critical initial amplitude, the amplitude of the wave becomes unbounded in a finite time. This, of course, suggests the formation of a shock. As noted in Ref. 1, the case $\omega_0 < 0$, which has no mechanical analog, is the case of primary interest and importance. The foregoing analytical treatment shows that in this case the amplitude of the wave either tends to λ_0 eventually or else becomes unbounded in a finite time. Furthermore, note that in this case if $a(0)$ and β_0 have the same sign, the amplitude of the wave always becomes unbounded in a finite time. Moreover, since $a(0)$ arises from the thermal noise, there are always some $a(0)$ with the same sign as β_0.

Let us now consider the behavior of a wave for which β_0 vanishes. In our earlier treatment of one-dimensional acceleration waves,[1] we noted that the vanishing of β_0 corresponded to a linear material. Indeed, while β_0 also vanishes identically here if the response of the material is linear, it may also vanish because of a combination of other factors even though the response of the material is nonlinear. For example, in a given material, once $\mathscr{E}$ and $\mathbf{F}$ are prescribed it may be possible to choose $\mathbf{n}$ [and hence $\mathbf{r}$ through Eq. (3.35)] in such a way that β_0 vanishes. In particular, it is easily verified in the relatively simple case of a purely transverse acceleration wave propagating in the direction of the applied electric field and of a principal axis of homogeneous deformation in an isotropic material that β_0 vanishes even though the response of the material is nonlinear. In this simplest but extremely important case, Eq. (4.33) has the solution

$$a(t) = a_0 e^{-\omega_0 t}, \tag{5.9}$$

which means that if $\omega_0 > 0$, $a(t)$ is a monotonically decreasing function of time and the amplitude of the wave decreases

as the wave traverses the material. On the other hand, if $\omega_0 < 0$, $a(t)$ is a monotonically increasing function of t and the amplitude of the wave increases without bound as the wave traverses the material. Of course, if $\omega_0 = 0$, then $a(t) = a_0$ so that the wave propagates at constant amplitude.

Let us recall from Eqs. (3.28), (4.28), and (4.34) that

$$\omega_0 = \frac{1}{2\rho U}\left(\zeta\mu^e L_i r_i - \nu\frac{\partial p^e}{\partial \mu^e}\Sigma_i r_i\right)n_j r_j. \tag{5.10}$$

In particular, Eq. (5.10) shows that ω_0 vanishes whenever $\mathbf{r}$ is orthogonal to $\mathbf{n}$. Thus, if β_0 vanishes, either because the response of the material is linear or because $\mathscr{E}$, $\mathbf{F}$, and $\mathbf{n}$ have appropriate values, then purely transverse acceleration waves will propagate at constant amplitude and not grow and, of course, purely transverse shocks will not form. On the other hand, in this very special case of purely transverse acceleration waves, suppose that β_0 does not vanish for plane wave propagation in a prescribed direction $\mathbf{n}$; then, Eq. (4.33) reduces to

$$\frac{\delta a}{\delta t} = \beta_0 a^2, \tag{5.11}$$

so that

$$a = a_0(1 - \beta_0 a_0 t)^{-1}. \tag{5.12}$$

Of course, the solution (5.12) has the same form as the corresponding solution for acceleration waves in nonheat conducting elastic media. Note that if a_0 and β_0 have the same sign, then a shock will form after a time

$$\bar{t}_\infty = 1/\beta_0 a_0, \tag{5.13}$$

and, as already noted, since $a(0)$ arises from the thermal noise, a shock will always form for nonzero β_0 when ω_0 is zero.

When β_0 vanishes and the acceleration wave is not *purely* transverse, it should be clear from the above discussion that the threshold condition, at which the amplitude $a(t)$ just begins to grow, may be defined by

$$\hat{\omega}_0 = 0, \tag{5.14}$$

where

$$\hat{\omega}_0 = \frac{1}{2\rho U}\left(\zeta\mu^e L_i r_i - \nu\frac{\partial p^e}{\partial \mu^e}\Sigma_i r_i\right). \tag{5.15}$$

A most important limiting form of Eq. (2.16) [or Eq. $(2.7)_6$ with Eqs. (2.12), $(2.8)_7$, $(2.9)_3$, and $(2.7)_5$] is

$$\mathscr{J}_K = -J\mu^e m_{KL}\mathscr{E}_L - JD^C_{KL}G_L, \tag{5.16}$$

where m_{KL} is the mobility tensor and D^C_{KL} is the diffusivity tensor, which may be written in the form

$$D^C_{KL} = -m_{KL}\partial p^e/\partial\mu^e, \tag{5.17}$$

and, of course, we can have $m_{KL} = m_{KL}(E_{KL}, \mathscr{E}_L)$. In this simple but important limiting case, in which the current is given by Eq. (5.16) with the mobility tensor m_{KL} and diffusivity tensor D^C_{KL} constant, from Eqs. (5.15), (4.18), (4.11), (4.13), (5.16), (2.21), and (5.17) and the fact that G_K vanishes ahead of the wave front, we obtain

$$\hat{\omega}_0 = -\frac{1}{2\rho U}\nu\frac{\partial p^e}{\partial \mu^e}\mu^e N_K J X_{K,i} r_i [N_L m_{LR}\mathscr{E}_R + U_N], \tag{5.18}$$

from which, with Eq. (5.14), we find that the threshold relation is given by

$$N_L m_{LR}\mathscr{E}_R + U_N = 0. \tag{5.19}$$

When the deformation is infinitesimal, we have

$$F_{rL} \approx \delta_{rL},\ \mathscr{E}_R \approx \delta_{Rj}E_j = E_R,\ U_N \approx U = U_0, \tag{5.20}$$

which enables us to write Eq. (5.19) in the form

$$N_L m_{LR}\mathscr{E}^T_R = N_L m_{LR}E^T_R = -U_0, \tag{5.21}$$

where it should be recalled that $\mathbf{N}$ (or $\mathbf{n}$) is the normal to the plane wave surface. Equation (5.21) is the generalization of the well-known relation for the threshold field obtained in Eq. (5.13) of Ref. 1 for the one dimensional case to the arbitrarily anisotropic three-dimensional case treated here for the restricted limiting form (5.16) of the current equation (2.16).

6. WEAK WAVES AND INDUCED DISCONTINUITIES

Following Coleman and Gurtin,[24] we define a wave of order N as follows:

A propagating singular surface Σ is a wave of order N if the field $\mathbf{y}(\mathbf{X},t)$ and its first $N-1$ partial derivatives with respect to $\mathbf{X}$ and t are continuous everywhere, but the Nth order partial derivatives suffer jump discontinuities at Σ, and are continuous functions everywhere else. In particular, we note that the case $N = 2$ represents an acceleration wave. If $N > 2$, the wave is said to be a weak wave.

Our object here is to study the propagation and growth of weak waves in a piezoelectric semiconducting material. It suffices to consider waves of order 3. We confine our attention to the study of plane waves and we assume that the material ahead of the wave front is in a state of homogeneous strain, is subject to a uniform electric field, and that the charge density of the free electronic fluid is uniform and constant prior to the arrival of the wavefront.

Since we are dealing with plane waves of order 3, from the compatibility conditions across Σ we have

$$\begin{aligned}
&[\dot{F}_{rL,K}] = [v_{r,LK}] = b_r N_L N_K,\\
&[F_{rL,KM}] = -U_N^{-1} b_r N_L N_K N_M,\\
&[\ddot{F}_{rL}] = [\dot{v}_{r,L}] = -U_N b_r N_L,\\
&[\ddot{v}_r] = U_N^2 b_r,\\
&[\ddot{F}_{rK,L}] = [\dot{v}_{r,KL}] = d_r N_K N_L,\ d_r = [N_M N_S \ddot{F}_{rS,M}].\\
&[\dot{v}_r] = U_N^2 d_r + 2U_N^2\frac{\delta b_r}{\delta t}.
\end{aligned} \tag{6.1}$$

It follows from Eqs. (2.11) and (3.16) in essentially the same way that Eq. (3.19) follows from Eq. (3.18) that $\mathscr{E}_{L,K}$ is continuous across Σ and since $\mathscr{E}_L$ is continuous across Σ also, from the kinematic condition of compatibility $\dot{\mathscr{E}}_L$ is continuous across Σ. From the geometric condition of compatibility, we have

$$\begin{aligned}
&[\dot{\mathscr{E}}_{L,K}] = \beta N_L N_K,\quad [\ddot{\mathscr{E}}_{L,K}] = \bar{\beta} N_L N_K,\\
&\bar{\beta} = -[N_R N_S \ddot{\varphi}_{,RS}].
\end{aligned} \tag{6.2}$$

In a similar manner Eqs. (2.13), $(3.8)_2$, and (4.15) imply that $G_{K,L}$ is continuous across Σ and since G_K is continuous

across Σ also, from the kinematic condition of compatibility $\dot{G}_K$ is continuous across Σ, but from the geometric condition of compatibility and Eq. (2.17), we have

$$[\ddot{G}_K] = \bar{\bar{\omega}} N_K, \quad \bar{\bar{\omega}} = [N_R \ddot{G}_R]. \tag{6.3}$$

Since for weak waves a_j vanishes, an immediate consequence of Eq. (4.27) is that

$$(\hat{Q}_{jp} - \rho_0 U_N^2 \delta_{jp}) b_p = 0, \tag{6.4}$$

so that the propagation condition for weak waves is precisely the same as that which governs the propagation of acceleration waves.

In order to obtain the differential equation which governs the evolutionary behavior of the amplitudes of weak waves, we differentiate Eq. (4.1) with respect to t holding $\mathbf{X}$ fixed, and on taking the jump across Σ in the resulting equation, employing Eqs. (3.34), (4.5), $(6.1)_2$, and (6.3), and recalling the steady uniform state ahead of the wavefront, we obtain

$$(A_{jKpL} N_K N_L - \rho_0 U_N^2 \delta_{jp}) d_\rho + \bar{\beta} L_j + \bar{\bar{\omega}} H_{jK} N_K - 2\rho_0 U_N^2 \frac{\delta b_j}{\delta t} = 0. \tag{6.5}$$

Similarly, taking the material time derivatives of Eqs. (4.10) and (4.19), taking the jumps across Σ in the resulting equations, employing Eqs. $(2.21)_1$, (3.34), (4.5), (4.8), (4.13), (4.14), (4.22), (6.1)–(6.3), and the compatibility conditions, and recalling the steady uniform state ahead of the wavefront, we find that

$$\bar{\beta} = \zeta L_i d_i - \zeta U_N J \mu^e X_{K,i} N_K b_i, \quad \bar{\bar{\omega}} = -U_N \nu \Sigma_i b_i. \tag{6.6}$$

Substituting from Eq. (6.6) into (6.5) and employing Eqs. (3.36) and (4.25), we obtain

$$(\hat{Q}_{jp} - \rho_0 U_N^2 \delta_{jp}) d_p = \alpha_{jp} b_p + 2\rho_0 U_N^2 \frac{\delta b_p}{\delta t}. \tag{6.7}$$

Let us write

$$b_j = U_N^2 c_j, \quad c_j = c r_j, \quad |\mathbf{r}| = 1, \tag{6.8}$$

where $\mathbf{r}$ is the unit eigenvector of Q_{jp} corresponding to the eigenvalue $\rho_0 U_N^2$. If we now contract Eq. (6.7) with r_j and employ Eqs. (3.5), $(3.8)_6$, (3.35), (3.36), (4.28), and (4.34), we find that the amplitude c of the third-order wave satisfies the equation

$$\frac{\delta c}{\delta t} = -\omega_0 c, \tag{6.9}$$

which admits the solution

$$c(t) = c_0 e^{-\omega_0 t}. \tag{6.10}$$

It is now clear that the evolutionary behavior of the amplitude of a weak plane wave is somewhat different from that of a plane acceleration wave which propagates in the same direction at the same speed. In particular, the evolutionary behavior of a weak wave is determined solely by the sign of ω_0 and is independent of the initial value of the amplitude of the wave. As we have noted earlier, ω_0 may be negative in certain important circumstances. When ω_0 is negative, the amplitude of the weak wave will increase without bound and become an acceleration wave as the wave traverses the material. This behavior should be contrasted with the manner in which weak waves behave in other media (see for example Ref. 6). However, if the weak wave is purely transverse, the amplitude will remain constant as the wave traverses the material in accordance with the relevant portion of the discussion in Sec. 5.

Let us now suppose that there exists a particular direction $\mathbf{n}$ in which three real plane acceleration waves may propagate. Let us denote the amplitudes, unit amplitude vectors, and speeds of propagation of these waves by $a^{(i)}$, $\mathbf{r}^{(i)}$, and $U_N^{(i)}$, $i = 1,2,3$, respectively. It follows from Eq. (4.32) that the amplitude $\mathbf{b}^{(i)}$ of the third-order discontinuity induced by the acceleration wave of amplitude $a^{(1)}$ is determined by the equations

$$(\hat{Q}_{jp} - \rho_0 U_N^{(1)2} \delta_{jp}) b_p^{(1)} = d_j^{(1)}, \tag{6.11}$$

with

$$d_j^{(1)} = 2\rho_0 U^{(1)2} r_j^{(1)} \frac{\delta a^{(1)}}{\delta t} + \bar{\alpha}_{jp}^{(1)} r_p^{(1)} a^{(1)} + \bar{\alpha}_{jpq}^{(1)} r_p^{(1)} r_q^{(1)} a^{(1)2}, \tag{6.12}$$

where $\bar{\alpha}_{jp}^{(1)}$ and $\bar{\alpha}_{jpq}^{(1)}$ are given by Eqs. (4.28) and (4.29), respectively, with U replaced by $U^{(1)} = B_n U_N^{(1)}$.

Now suppose that the eigenvectors of $\hat{Q}_{ij}(\mathbf{n})$ are distinct so that the vectors $\mathbf{r}^{(i)}$ form an orthogonal triad. We may now write $\mathbf{b}^{(1)}$ in the form

$$\mathbf{b}^{(1)} = \sum_{\alpha=1}^{3} b_\alpha^{(1)} \mathbf{r}^{(\alpha)}. \tag{6.13}$$

Since, as a consequence of Eq. (4.33), $d_j^{(1)}$ is orthogonal to $r_j^{(1)}$, even though $U_N^{(1)2}$ is a root of Eq. (3.37), Eqs. (6.11) are consistent but do not determine the component of $\mathbf{b}^{(1)}$ in the direction of $\mathbf{r}^{(1)}$ uniquely. Nevertheless, when Eqs. (6.11) are contracted successively with $r_j^{(2)}$ and $r_j^{(3)}$ and Eq. (3.37) is employed, it follows that

$$\rho_0 (U_N^{(2)2} - U_N^{(1)2}) b_2^{(1)} = r_j^{(2)} d_j^{(1)},$$
$$\rho_0 (U_N^{(3)2} - U_N^{(1)2}) b_3^{(1)} = r_j^{(3)} d_j^{(1)}. \tag{6.14}$$

Of course, Eqs. (6.14) determine uniquely the components of the induced discontinuity in the two mutually orthogonal directions which are also orthogonal to the direction of the amplitude vector of the primary acceleration wave.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Research under Contract No. N00014-76-C-0368 and the National Science Foundation under Grant No. ENG 72-04223.

[1] M.F. McCarthy and H.F. Tiersten, "One-dimensional acceleration waves and acoustoelectric domains in piezoelectric semiconductors," J. Appl. Phys. **47**, 3389 (1976).
[2] H.G. deLorenzi and H.F. Tiersten, "On the interaction of the electromagnetic field with heat conducting deformable semiconductors," J. Math. Phys. **16**, 938 (1975).
[3] M.F. McCarthy and H.F. Tiersten, "On Integral Forms of the Balance Laws for Deformable Semiconductors," Arch. Rational Mech. Anal. **68**, 27 (1978).

[4]B.D. Coleman, M.E. Gurtin and I. Herrera R., "Waves in Materials with Memory, I. The Velocity of One-Dimensional Acceleration Waves," Arch. Ratl. Mech. Anal. **19**, 1 (1965).
[5]B.D. Coleman and M.E. Gurtin, "Waves in Materials with Memory, II. On the Growth and Decay of One-Dimensional Acceleration Waves," Arch. Ratl. Mech. Anal. **19**, 239 (1965).
[6]B.D. Coleman and M.E. Gurtin, "Waves in Materials with Memory, III. Thermodynamic Influences on the Growth and Decay of Acceleration Waves," Arch. Ratl. Mech. Anal. **19**, 266 (1965).
[7]E. Varley, "Acceleration Fronts in Viscoelastic Materials," Arch. Rational Mech. Anal. **19**, 215 (1965).
[8]M.F. McCarthy, "The Growth of Magnetoelastic Waves in a Cauchy Elastic Material in a Magnetic Field," Arch. Rational Mech. Anal. **23**, 191 (1966).
[9]P.J. Chen, "Thermodynamic Influences on the Propagation and Growth of Acceleration Waves in Elastic Materials," Arch. Rational Mech. Anal. **31**, 228 (1968).
[10]M.F. McCarthy, "Thermodynamic Influences on the Propagation of Waves in Electroelastic Materials," Int. J. Eng. Sci. **11**, 1301 (1973).
[11]P.J. Chen, "Growth and Decay of Waves in Solids," in *Encyclopedia of Physics*, edited by C. Truesdell (Springer, Berlin, 1973), Vol. VI, a/3, pp. 303–402.
[12]M.F. McCarthy, "Singular Surfaces and Waves," *Continuum Physics*, edited by A.C. Eringen (Academic, New York, 1975), Vol. II, Chap. 2.
[13]Such a definition has been used in this manner before by R.N. Thurston, "Waves in solids," in *Encyclopedia of Physics*, edited by C. Truesdell (Springer, Berlin, 1974), Vol. VIa/4.
[14]Reference 12, Sec. 2.2, Eq. (2.2.25).
[15]Reference 12, Sec. 2.2, Eq. (2.2.27).
[16]Reference 12, Sec. 2.3, Eqs. (2.3.2) and (2.3.3).
[17]C. Truesdell, "General and Exact Theory of Waves in Finite Elastic Strain," Arch. Ratl. Mech. Anal. **8**, 263 (1961).
[18]C. Truesdell, "Existence of Longitudinal Waves," J. Acoust. Soc. Am. **40**, 729 (1966).
[19]C. Truesdell and W. Noll, "The Nonlinear Field Theories of Mechanics," in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1965), Vol. III/3, pp. 1–602.
[20]C.C. Wang and C. Truesdell, *Introduction to Rational Elasticity* (Noordhoff, Leyden, 1973).
[21]P. Chadwick and P.K. Currie, "On the Existence of Transverse Acceleration Waves," Math. Proc. Cambridge Philos. Soc. **77**, 405 (1975).
[22]Reference 12, Sec. 2.3, Eqs. (2.3.4) and (2.3.5).
[23]Reference 12, Sec. 2.2.
[24]B.D. Coleman and M.E. Gurtin, "Waves in Materials with Memory, IV. Thermodynamics and the Velocity of General Acceleration Waves," Arch. Ratl. Mech. Anal. **19**, 317 (1965).

Singular perturbations approach to the limit cycle and global patterns in a nonlinear diffusion-reaction problem with autocatalysis and saturation law

L. L. Bonilla and M. G. Velarde[a)]
Departamento de Física, C-3, Universidad Autónoma de Madrid, Cantoblanco (Madrid), Spain

(Received 25 July 1978)

A two-time scales method (singular perturbations) is used to construct the limit cycle and the global nonuniform steady patterns that appear in a diffusion-reaction process with autocatalysis and the Langmuir–Hinshelwood (Michaelis–Menten, Holling) saturation law. The stability of these nonlinear structures is also established in this paper.

1. INTRODUCTION

In two previous papers Ibáñez and Velarde[1,2] have discussed the existence and the stability properties of the multiple nonuniform steady states that arise in a diffusion-reaction process of eventual relevance to heterogeneous catalysis (with catalytic wires, say) and surface reaction processes, to enzyme-controlled reactions in biophysical chemistry and to predator–prey competition schemes in ecological context.[3-5] The authors restricted their study, however, to the stationary solutions in the limiting approximation of large scale separation between the two relevant diffusion parameters of the problem. In the present paper we extend the previous work by considering a more general case with two arbitrary diffusion constants. We construct and delineate the stability properties of global two-component nonuniform patterns that are asymptotically expected in an initially homogeneous one-dimensional reactor. We also construct the time-periodic solutions (limit cycle and unstable orbit) that the diffusion-reaction process can exhibit. The model consideration refers to the following process[3]:

$$A \xrightarrow{k_1} Y, \tag{1.1a}$$

$$X + Y \xrightarrow{k_2} 2Y \quad \text{(autocatalysis)}, \tag{1.1b}$$

$$X \xrightarrow{S(k_3,k_4)} P \quad \text{(saturation law)}, \tag{1.1c}$$

in which all three steps are taken irreversible, A, P, X, and Y denote reactants and the system is considered open to in- and out-flow of A and P but the concentration of these two products is kept uniform all throughout the reactor. For simplicity we shall restrict consideration here to one-dimensional reactors only. $S(k_3,k_4)$ accounts for the Langmuir–Hinshelwood law in heterogeneous catalysis and adsorption at surfaces, the Michaelis–Menten law in enzyme-controlled processes, and the Holling law in ecology.

Disregarding convective phenomena and considering isothermal processes only, the scheme (1) can be described by the following set of nonlinear partial differential equations:

$$\frac{\partial [X]}{\partial t} = k_2[X][Y] - k_3 \frac{[X]}{1 + k_4[X]} + D_{[X]} \frac{\partial^2 [X]}{\partial r^2}, \tag{1.2a}$$

$$\frac{\partial [Y]}{\partial t} = k_1[A] - k_2[X][Y] + D_{[Y]} \frac{\partial^2 [Y]}{\partial r^2}, \tag{1.2b}$$

where bracketed quantities denote concentration of reactants; $D_{[X]}$ and $D_{[Y]}$ are respectively the two diffusion parameters in the system that we shall take as constants. k_i ($i = 1,2,3$) are reaction rate constants and k_4 accounts for the strength of the saturation law.

For universality of the description that follows we shall take convenient "scales" and define the following dimensionless quantities:

$$\left\{ X = \frac{k_2}{k_3}[X], \quad Y = \frac{k_2}{k_3}[Y], \quad A = \frac{k_1 k_2}{k_3}[A], \right.$$
$$q = \frac{k_3 k_4}{k_2}, \quad p = \frac{k_1}{k_3}, \quad D_X = D_{[X]}/k_3 L^2,$$
$$\left. D_Y = D_{[Y]}/k_3 L^2, \quad \tilde{r} = \frac{r}{L}, \quad \tilde{t} = k_3 t \right\}.$$

L denotes the length of the one-dimensional reactor. As no confusion is to be expected we shall for convenience drop the tilde on r and t. Thus in dimensionless form the system (1.2) reduces to the following two-variable problem:

$$\frac{\partial X}{\partial t} = XY - \frac{X}{1 + qX} + D_X \frac{\partial^2 X}{\partial r^2}, \tag{1.3a}$$

$$\frac{\partial Y}{\partial t} = A - XY + D_Y \frac{\partial^2 Y}{\partial r^2}, \tag{1.3b}$$

in which A is a constant value. Obviously X and Y ought to satisfy initial conditions (i.c.) and boundary conditions (b.c.) yet to be specified below.

The system (1.3) possesses a trivial fixed point (i.e., a steady homogeneous solution)

$$X_s = A/(1 - qA), \quad Y_s = 1 - qA, \tag{1.4}$$

whose actual existence depends on the b.c. For simplicity we shall consider in the following that the concentrations are fixed at the boundaries (Dirichlet problem) such that

a)Author to whom all correspondence should be addressed.

 0022-2488/79/122692-12$01.00

$$X(r=0,t)=X(r=1,t)=X_s, \tag{1.5a}$$

$$Y(r=0,t)=Y(r=1,t)=Y_s. \tag{1.5b}$$

This particular choice eliminates the appearence of boundary layers at extreme ends of the reactor where the homogeneous solution (1.4) should adjust to the imposed values.

In Sec. 2 we study the stability of the uniform steady solution (1.4). In Sec. 3 the construction of global two-component nonuniform patterns is obtained starting with a uniform distribution of reactants in the vessel. Section 4 deals with the construction in the same limit of time-periodic solutions (limit cycle and unstable orbit) of (1.3). A discussion is also given about the influence of diffusion upon the limit cycle operation in the reactor. For general background see the recent monographs by Prigogine and collaborators[6,7] and the book on synergetics by Haken.[8]

2. STABILITY OF THE STEADY UNIFORM DISTRIBUTION OF REACTANTS IN THE PRESENCE OF DIFFUSION

Initial conditions will only be formally introduced in this section as we shall consider here possible two-component global nonuniform steady states in the reactor. For convenience in the analysis that follows, and to be able to compare with the results found in Refs. 1 and 2, we introduce $D \equiv D_Y$, and $\theta = D_X/D_Y$. Thus our problem can be recast in the compact form

$$\frac{\partial u}{\partial t} = L(\gamma)u + N(\gamma,u), \tag{2.1}$$

where, $u = \binom{x}{y}$, $\gamma = \{q,A,\theta,D\}$, and $x = X - X_s$, $y = Y - Y_s$, in which X_s, Y_s correspond to (1.4) and $X = X(r,t)$, $Y = Y(r,t)$ are the actual values of the concentration of the intermediate reactants. We shall consider asymptotic solutions to (2.1), i.e., for t going to infinity. The finding of nontrivial ($u \not\equiv 0$) solutions to (2.1) with b.c. $u(0,t) = u(1,t) = 0$, corresponds to an instability of the steady uniform distribution (1.4) of reactants in the vessel, and thus to the appearance (bifurcation) of secondary patterns in (1.3). The parameter γ introduced above accounts for either the externally controlled quantities or the diffusion constants. The following definitions have also been used:

$$L(\gamma) \equiv \begin{pmatrix} qA(1-qA) + \theta D\dfrac{\partial^2}{\partial r^2} & \dfrac{A}{(1-qA)} \\ -(1-qA) & \dfrac{-A}{1-qA} + D\dfrac{\partial^2}{\partial r^2} \end{pmatrix}, \tag{2.2}$$

$$N(\gamma,u) \equiv \begin{pmatrix} xy + \displaystyle\sum_{n=2}^{\infty} (-)^n q^{n-1}(1-qA)^{n+1}x^n \\ -xy \end{pmatrix}. \tag{2.3}$$

A sufficient condition for instability of (1.5) is that $u = 0$ be unstable to infinitesimal disturbances, and thus that the linearized approximation to (2.1),

$$\left[\frac{\partial}{\partial t} - L(\gamma)\right]u = 0, \tag{2.4}$$

would have a nontrivial solution for the b.c. specified above. Formal solutions to (2.4) are

$$u(r,t) = \Xi(r)e^{\sigma t}, \tag{2.5}$$

where $\Xi(r) \equiv \binom{\xi(r)}{\eta(r)}$ corresponds to the vector solution of the time-independent part in (2.2). Insertion of (2.5) in (2.4) yields the following eigenvalue problem (with parameter σ),

$$[L(\gamma) - \sigma I]\Xi(r) = 0, \tag{2.6}$$

in which I denotes the identity matrix in dimension two. For the time being we assume that all eigenvalues in (2.6) are simple. Later on this assumption will be justified. Thus the general solution to (2.4) can be expressed as a linear combination of (2.5) forms

$$u(r,t) = \sum_{n=1}^{\infty} c_n e^{\sigma_n t}\Xi_n(r) \tag{2.7}$$

of which we shall indeed take the real part only. σ_n or σ is the time constant that determines stability. It suffices for a single σ_n to have $\mathrm{Re}\sigma_n > 0$ for the solution $u = 0$ to be unstable.

With the above specified Dirichlet b.c. and using the eigenfunctions of the Laplacian we can write

$$\xi_n(r) = \sin n\pi r, \qquad \eta_n(r) = M_n \sin n\pi r.$$

Thus Eq. (2.6) yields the following characteristic equation

$$\sigma_n^2 - T(\gamma,n)\sigma_n + D(\gamma,n) = 0, \tag{2.8a}$$

if a nontrivial solution is to exist. Here

$$T(\gamma,n) \equiv qA(1-qA) - \frac{A}{1-qA} - n^2\pi^2 D(1+\theta), \tag{2.8b}$$

$$D(\gamma,n) = A(1-qA) - n^2\pi^2 D\left[qA(1-qA) - \frac{A\theta}{1-qA}\right] + (n^2\pi^2 D)^2\theta. \tag{2.8c}$$

To every n there is associated a couple of roots in (2.8a) that may be real or complex conjugates depending on the values given to a and A. For later convenience we shall call these two eigenvalues $\sigma_n^{\pm}$ according to the following convention,

$$2\sigma_n^{\pm} = qA(1-qA) - \frac{A}{1-qA} - n^2\pi^2 D(1+\theta) \pm \left\{\left[qA(1-qA) + \frac{A}{1-qA} + n^2\pi^2 D(1-\theta)\right]^2 - 4A\right\}^{1/2}. \tag{2.9}$$

It is clear that $\mathrm{Re}\sigma_n^+ \geqslant \mathrm{Re}\sigma_n^-$ though the eigenvalue with highest real part is yet to be determined. From the values $\sigma_n^{\pm}$ we obtain the eigenfunctions by solving

$$\sigma_n^{\pm} - qA(1-qA) + n^2\pi^2 D\theta - \frac{A}{1-qA}M_n^{\pm} = 0. \tag{2.10}$$

Given q, θ, and D there are two possible modes of instability at some critical value A_c (call $\gamma_c = [A_c,q,\theta,D]$). We shall separately study them. They are:

(i) At $\gamma = \gamma_c$ an eigenvalue $\sigma_{n_c}^{\pm}$ crosses the imaginary axis with nonvanishing imaginary part (Hopf bifurcation). This case corresponds to

$$T(\gamma_c,n_c) = 0, \tag{2.11a}$$

$$D(\gamma_c,n_c) > 0, \tag{2.11b}$$

$$T(\gamma_c,n\neq n_c) < 0, \tag{2.11c}$$

$$D(\gamma_c,n\neq n_c) > 0, \tag{2.11d}$$

with all other eigenvalues $\sigma_{n\neq n_c}$ having negative real parts and thus decaying on the time scale $t\to\infty$. From (2.11a) and (2.11c) follows that $n_c = 1$. In the plane (q,A) the locus

$$qA_c(1-qA_c) - \frac{A_c}{1-qA_c} = \pi^2 D(1+\theta) \tag{2.12}$$

corresponds to neutral stability ($\mathrm{Re}\sigma_{n_c}^{+} = 0$).

(ii) At $\gamma = \gamma_c$ the only eigenvalue n_c that crosses the imaginary axis, from negative to positive has vanishing imaginary part. This corresponds to

$$T(\gamma_c,n_c) < 0, \tag{2.13a}$$

$$D(\gamma_c,n_c) = 0, \tag{2.13b}$$

$$T(\gamma_c,n\neq n_c) < 0, \tag{2.13c}$$

$$D(\gamma_c,n\neq n_c) > 0, \tag{2.13d}$$

From (2.13b) and (2.13d) it follows that the eigenvalue that yields instability corresponds to the minimum of D. This condition together with (2.8c) gives a nonvanishing n_c,

$$n_c = \|\pi^{-1}D^{-1/2}\theta^{-1/4}A_c^{1/4}(1-qA_c)^{1/4}\|, \tag{2.14}$$

where $\|x\|$ denotes the integer part of x. The locus

$$qA_c(1-qA_c) + \frac{A_c\theta}{1-qA_c} = 2[\theta A_c(1-qA_c)]^{1/2} \tag{2.15}$$

corresponds to neutrally stable states.

It is to be noted that inserting (2.15) in (2.13a) the following inequality holds

$$\left(\frac{\theta-1}{2\theta}\right)\left[qA_c(1-qA_c) - \frac{A_c\theta}{1-qA_c}\right] < 0, \tag{2.16}$$

and from the sufficient condition of minimum to D

$$\theta < 1, \quad \text{i.e.,} \quad D_X < D_Y, \tag{2.17}$$

which comes with the sufficient condition of instability of the trivial solution. Figure 1 depicts for some specific values of γ the (linear) stability diagram of (2.1) in the neighborhood of $u = 0$. It should be noted that with respect to the stirred reactor in the present case of an unstirred vessel, diffusion plays either a stabilizing or a destabilizing role, depending on the conditions of operation, i.e., on the values given to A, D, θ, and q. All these results agree well with previous results obtained by Velarde *et al.*[1,2,5,9]

Before closing this section we shall introduce some definitions and remarks and recall a number of properties of the linearized operator $L(\gamma)$ to be used in the subsequent sections:

(i) If $\sigma_n^{\pm}$ are complex eigenvalues, then $M_n^{-} = M_n^{+*}$ in which the *star* denotes complex conjugation.

(ii) Let F be the space of analytic functions $u(r) = \binom{x(r)}{y(r)}$ such that $u(0) = u(1) = 0$. Then

$$\langle u \mid \bar{u} \rangle = \int_0^1 \{x^*(r)\bar{x}(r) + y^*(r)\bar{y}(r)\}\, dr \tag{2.18}$$

is a scalar product in F.

(iii) From (2.9) and (2.10) it follows that

$$M_n^{+}M_n^{-} = \frac{(1-qA)^2}{A}. \tag{2.19}$$

(iv) Let $\hat{L}(\gamma)$ be the adjoint operator to $L(\gamma)$ and

$$\hat{\Xi}_n^{\pm}(r) = \begin{pmatrix} \sin n\pi r \\ N_n^{\pm}\sin n\pi r \end{pmatrix}$$

the eigenfunctions of $\hat{L}(\gamma)$ with same Dirichlet b.c. It follows

$$N_n^{\pm} = -\frac{A}{(1-qA)^2} M_n^{\pm *}. \tag{2.20}$$

It is clear that whenever $\sigma_n^{\pm}$ are real both $M_n^{\pm}$ and $N_n^{\pm}$ are also real valued.

(v)

$$\langle \hat{\Xi}_m^{\pm} \mid \Xi_n^{\pm} \rangle = \frac{1}{2}\left(1 - \frac{A}{(1-qA)^2} M_n^{\pm} M_m^{\pm}\right)\delta_{n,m}, \tag{2.21}$$

$$\langle \hat{\Xi}_n^{\mp} \mid \Xi_n^{\pm} \rangle = 0, \tag{2.22}$$

Thus the set $\Xi_n^{\pm}(r)$ defines an orthogonal complete set (basis) in F, and for an arbitrary function $f(r)$ belonging to F the following expansion holds,

$$f(r) = \sum_{n=1}^{\infty} (\alpha_n^{+}\Xi_n^{+}(r) + \alpha_n^{-}\Xi_n^{-}(r)), \tag{2.23}$$

where

$$\alpha_n^{\pm} = \frac{2\langle \hat{\Xi}_n^{\pm} \mid f \rangle}{1 - [A/(1-qA)^2]M_n^{\pm 2}}. \tag{2.24}$$

(iv) At the critical point, $A = A_c$ if the eigenvalue with vanishing real part is real, then

$$M_{n_c}^{+} = -\tfrac{1}{2}[\theta + q(1-qA_c)^2]. \tag{2.25}$$

3. GLOBAL NONUNIFORM STEADY PATTERNS

In this section we shall construct for Dirichlet b.c. the nonuniform steady solution that branches at $A = A_c$, in region III of Figure 1. We shall define two time-scales and develop a convenient asymptotic expansion approach.

Let

$$x(r,0) = h(r,\epsilon); \quad h_\epsilon(r,0) = \left(\frac{\partial h(r,\epsilon)}{\partial \epsilon}\right)_{\epsilon=0}, \tag{3.1a}$$

$$y(r,0) = g(r,\epsilon); \quad g_\epsilon(r,0) = \left(\frac{\partial g}{\partial \epsilon}\right)_{\epsilon=0}, \tag{3.1b}$$

$$h(r,0) = g(r,0) = 0, \tag{3.1c}$$

be some given, though rather arbitrary, initial data which are

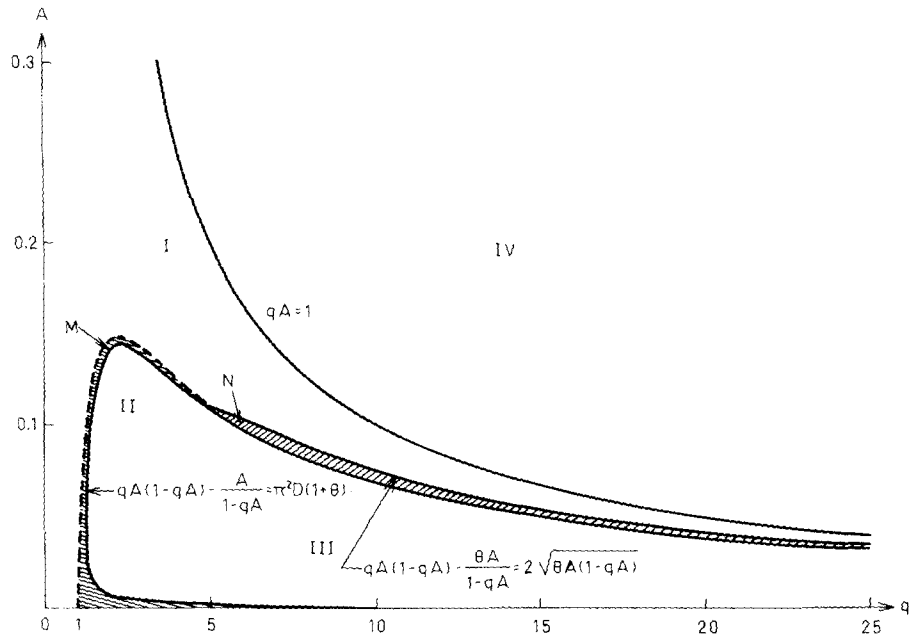

FIG. 1. Stability portrait of (1.3) in the neighborhood of the homogeneous steady state (1.4). Region I is of stability. Region IV corresponds to unphysical states of negative concentrations ($qA > 1$). Region II-III contains the unstable zone: on the left, along M [Eq. (2.12)] there is bifurcation to limit cycle behavior whereas on the right, along N [Eq. (2.15)] spatial dissipative structures are expected.

perturbations upon the trivial fixed point, $x = y = 0$. With the new unknown ϵ we denote a smallness parameter and we shall seek of bifurcating solutions to (2.1) in an ϵ-neighborhood of A_c. The following time-scales will be used: (i) a *fast scale* $\tilde{t} = t$ which goes along the trajectory of $x(r,t)$, $y(r,t)$, $t \geqslant 0$, and (ii) a *slow scale* $\tau = [A(\epsilon) - A_c]t$, which is an *ad hoc* scale to measure the distance from the actual concentration profiles to the asymptotic state eventually attained at $\tilde{t} \to \infty$ (and $\tau \to \infty$). We now define the following expansions,[10]

$$x(r,t,\tau) \sim \sum_{i=1}^{\infty} \epsilon^i x_i(r,t,\tau), \tag{3.2a}$$

$$y(r,t,\tau) \sim \sum_{i=1}^{\infty} \epsilon^i y_i(r,t,\tau), \tag{3.2b}$$

together with the corresponding expansions for the b.c.

$$x_j(0,t,\tau) = x_j(1,t,\tau) = y_j(0,t,\tau) = y_j(1,t,\tau) = 0, \tag{3.2c}$$

and i.c.

$$x_j(r,0,0) = \frac{1}{j!}\frac{\partial^j h(r,0)}{\partial \epsilon^j}, \tag{3.2d}$$

$$y_j(r,0,0) = \frac{1}{j!}\frac{\partial^j g(r,0)}{\partial \epsilon^j}. \tag{3.2e}$$

We shall also assume that A is analytic in ϵ in the neighborhood of A_c,

$$A(\epsilon) = A_c + A'(0)\epsilon + \tfrac{1}{2}A''(0)\epsilon^2 + o(\epsilon^3). \tag{3.3}$$

Thus the following formal expansions hold for the linear and nonlinear operators in (2.1):

$$L(\gamma) = \begin{pmatrix} qA_c(1-qA_c) + \theta D\dfrac{\partial^2}{\partial r^2} & A_c(1-qA_c)^{-1} \\ -(1-qA_c) & -A_c(1-qA_c)^{-1} + D\dfrac{\partial^2}{\partial r^2} \end{pmatrix} + \epsilon \begin{pmatrix} (1-2qA_c)qA'(0) & A'(0)(1-qA_c)^{-2} \\ qA'(0) & -A'(0)(1-qA_c)^{-2} \end{pmatrix}$$

$$+ \epsilon^2 \begin{pmatrix} \tfrac{1}{2}(1-2qA_c)qA''(0) + 2q^2A'^2(0) & \dfrac{(1-qA_c)A''(0) + 2qA'^2(0)}{2(1-qA_c)^3} \\ \tfrac{1}{2}qA''(0) & \dfrac{-(1-qA_c)A''(0) + 2qA'^2(0)}{2(1-qA_c)^3} \end{pmatrix} + o(\epsilon^3) \tag{3.4}$$

and

$$N(\gamma,u) = \epsilon^2 \begin{pmatrix} x_1y_1 + q(1-qA_c)^3x_1^2 \\ -x_1y_1 \end{pmatrix} + \epsilon^3 \begin{pmatrix} x_1y_2 + x_2y_1 + 2q(1-qA_c)^3x_1x_2 - 3q^2A'(0)(1-qA_c)^2x_1^2 - q^2(1-qA_c)^4x_1^3 \\ -x_1y_2 - x_2y_1 \end{pmatrix}$$

$$+ o(\epsilon^4). \tag{3.5}$$

The following formal differentiation applies,

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \tilde{t}} + [A'(0)\epsilon + \tfrac{1}{2}A''(0)\epsilon^2 + o(\epsilon^3)]\frac{\partial}{\partial \tau}, \tag{3.6}$$

and as in the following no confusion is expected we shall drop the tilde ($\sim$) on the time variable t.

Inserting the above expansion in Eq. (2.1) and equating the same powers in ϵ the following hierachy of linear equations is generated:

$$\mathcal{L}u_1 \equiv \begin{pmatrix} \frac{\partial}{\partial t} - qA_c(1-qA_c) - \theta D\frac{\partial^2}{\partial r^2} & -A_c(1-qA_c)^{-1} \\ (1-qA_c) & \frac{\partial}{\partial t} + A_c(1-qA_c)^{-1} - D\frac{\partial^2}{\partial r^2} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = 0, \tag{3.7a,b}$$

$$\mathcal{L}u_2 = \begin{pmatrix} -A'(0)\frac{\partial x_1}{\partial \tau} + qA'(0)(1-2qA_c)x_1 + A'(0)(1-qA_c)^{-2}y_1 + x_1y_1 + q(1-qA_c)^3x_1^2 \\ -A'(0)\frac{\partial y_1}{\partial \tau} + qA'(0)x_1 - A'(0)(1-qA_c)^{-2}y_1 - x_1y_1 \end{pmatrix}, \tag{3.8}$$

$$\mathcal{L}u_3 = \begin{pmatrix} -A'(0)\frac{\partial x_2}{\partial \tau} - \tfrac{1}{2}A''(0)\frac{\partial x_1}{\partial \tau} + q(1-2qA_c)A'(0)x_2 + [\tfrac{1}{2}qA''(0)(1-2qA_c) + qA'(0)^2]x_1 \\ + A'(0)(1-qA_c)^{-2}y_2 + [qA'(0)^2 + \tfrac{1}{2}A''(0)(1-qA_c)](1-qA_c)^{-3}y_1 + x_1y_2 + x_2y_1 + 2q(1-qA_c)^3x_1x_2 \\ - 3(1-qA_c)^2q^2A'(0)x_1^2 - q^2(1-qA_c)^4x_1^3 \\ -A'(0)\frac{\partial y_2}{\partial \tau} - \tfrac{1}{2}A''(0)\frac{\partial y_1}{\partial \tau} - A'(0)(1-qA_c)^{-2}y_2 - [qA'(0)^2 + \tfrac{1}{2}A''(0)(1-qA_c)] \\ \times(1-qA_c)^{-3}y_1 + qA'(0)x_2 + \tfrac{1}{2}qA''(0)x_1 - x_1y_2 - x_2y_1 \end{pmatrix} \tag{3.9}$$

...

The formal solution to (3.7) is

$$u_1(r,t,\tau) \equiv \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \mathrm{Re}\sum_{n=1}^{\infty}\{c_n^+(\tau)e^{\sigma_n^+ t}\Xi_n^+(r) + c_n^-(\tau)e^{\sigma_n^- t}\Xi_n^-(r)\}, \tag{3.10}$$

in which the dominant eigenvalue is $\sigma_{n_c}^+ = 0$, all other decay exponentially with t. Thus without loss of generality we can replace their respective coefficients $c_n^{\pm}(\tau)$ with their initial values [with the exception indeed of $c_{n_c}^+(\tau)$]. Thus (3.10) reduces to

$$\begin{aligned} u_1(r,t,\tau) &= c_{n_c}^+(\tau)\Xi_{n_c}^+(r) + c_{n_c}^-(0)e^{\sigma_{n_c}^- t}\Xi_{n_c}^-(r) + \mathrm{Re}\sum_{n\neq n_c}^{\infty}[c_n^{\pm}(0)e^{\sigma_n^- t}\Xi_n^{\pm}(r)] \\ &= c_{n_c}^+(\tau)\Xi_{n_c}^+(r) + (\text{e.d.t}), \end{aligned} \tag{3.11}$$

where (e.d.t) denotes "exponentially decaying terms."

The coefficients $c_n^{\pm}(0)$ can be obtained using (2.23) and (3.2). We have

$$c_n^{\pm}(0) = 2\left\langle \hat{\Xi}^{\pm} \middle| \begin{pmatrix} h_\epsilon & (r,0) \\ g_\epsilon & (r,0) \end{pmatrix} \right\rangle [1 - A_c(1-qA_c)^{-2}(M_n^{\pm 2})]^{-1}$$

$$= \frac{2\int_0^1 \{h_\epsilon(r,0) - A_c(1-qA_c)^{-2}M_n^{\pm}\, g_\epsilon(r,0)\}\sin n\pi r\, dr}{1 - A_c(1-qA_c)^{-2}(M_n^{\pm})^2}. \tag{3.12}$$

Thus all $c_n^{\pm}(0)$ are directly expressed in terms of the i.c., namely the arbitrary perturbation given to the trivial uniform solution to (2.1). The coefficient $c_{n_c}^{\pm}(\tau)$ is obtained by means of the ϵ^2 equation in the hierarchy (3.8), by using Fredholm's alternative. To later take the limit $t\to\infty$ it is convenient to introduce now the following average

$$\langle\langle \hat{\Xi}^{+}_{n_c} | f\rangle\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle \hat{\Xi}^{+}_{n_c} | f\rangle\, dt \tag{3.13}$$

for an arbitrary function in F. With the definition (3.13) all products of $\hat{\Xi}^{+}_{n_c}$ with (e.d.t.) vanish. We have

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T dt\int_0^1 dr\Bigg\{\sin^2 n_c\pi r\Bigg[-A'(0)\frac{dc^{+}_{n_c}}{d\tau}\left(1-\frac{A_c M^{+2}_{n_c}}{(1-qA_c)^2}\right)+A'(0)c^{+}_{n_c}\Bigg[q(1-2qA_c)+\frac{M^{+}_{n_c}}{(1-qA_c)^2}-\frac{qA_c M^{+}_{n_c}}{(1-qA_c)^2}$$
$$+\frac{A_c}{(1-qA_c)^4}M^{+2}_{n_c}\Bigg]\Bigg]+c^{+2}_{n_c}\sin^3 n_c\pi r\left[M^{+}_{n_c}+q(1-qA_c)^3+\frac{A_c M^{+2}_{n_c}}{(1-qA_c)^2}\right]\Bigg\}=0. \tag{3.14}$$

we recall that

$$\int_0^1 \sin^2 n_c\pi r\, dr = \tfrac{1}{2}, \tag{3.15}$$

and

$$\int_0^1 \sin^3 n_c\pi r\, dr = \begin{cases} 0 & \text{if } n_c \text{ is even} \\ \dfrac{4}{3n_c\pi} & \text{if } n_c \text{ is odd.}\end{cases} \tag{3.16}$$

Thus, from (3.14) it follows

$$A'(0)[1-A_c M^{+2}_{n_c}(1-qA_c)^{-2}]\frac{dc^{+}_{n_c}}{d\tau} = A'(0)c^{+}_{n_c}[q(1-2qA_c)+M^{+}_{n_c}(1-qA_c)^{-1}+A_c M^{+2}_{n_c}(1-qA_c)^{-2}]$$
$$+\frac{8c^{+2}_{n_c}}{3n_c\pi}[M^{+}_{n_c}+A_c(1-qA_c)^{-2}M^{+2}_{n_c}+q(1-qA_c)^{-3}], \tag{3.17}$$

where in accordance with (3.16) the second term in the r.h.s. of (3.17) is vanishing for n_c even.

Using now (2.24) and (2.14) in (3.17), we obtain

$$\tfrac{1}{2}A'(0)(1-\theta)\frac{dc^{+}_{n_c}}{d\tau}$$
$$= \frac{A'(0)}{4}c^{+}_{n_c}[q(1-3qA_c)+(1+qA_c)(1-qA_c)^{-2}\theta]$$
$$+c^{+2}_{n_c}\left\langle \hat{\Xi}^{+}_{n_c}\left|\left[\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2}N(\gamma,u)\right]_{\epsilon=0}\right\rangle\right., \tag{3.18}$$

where

$$\left\langle \hat{\Xi}^{+}_{n_c}\left|\left[\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2}N(\gamma,u)\right]_{\epsilon=0}\right\rangle\right.$$
$$= \begin{cases} \dfrac{q(1-qA_c)^2(\frac{1}{2}-qA_c)+\theta/2}{\frac{3}{4}n_c\pi} & \text{for } n_c \text{ odd,} \\ 0 & \text{for } n_c \text{ even.}\end{cases} \tag{3.19}$$

Thus we have two cases:

(i) When n_c is even, or odd but $q(1-qA_c)^2(\frac{1}{2}-qA_c)+\theta/2$ vanishes, then $A'(0)$ *must* be vanishing. For if $A'(0)$ is not vanishing it would be exponentially growing on time, and thus it would be a fast variable, contrary to hypothesis. Thus to obtain $c^{+}_{n_c}(\tau)$ the study of the ϵ^3 equation of the hierarchy (3.9) is needed. This case will be discussed further below.

(ii) When n_c is odd and $q(1-qA_c)^2(\frac{1}{2}-qA_c)+\theta/2 \neq 0$, Eq. (3.18) yields

$$\frac{dc^{+}_{n_c}}{d\tau} = \nu\left[1-\frac{c^{+}_{n_c}}{c^{+}_{n_c}(\infty)}\right]c^{+}_{n_c}, \tag{3.20a}$$

where

$$\nu = \frac{q(1-3qA_c)+[(1+qA_c)/(1-qA_c)^2]\theta}{2(1-\theta)} \tag{3.20b}$$

and

$$c^{+}_{n_c}(\infty) = -\frac{16A'(0)}{3n_c\pi}\times\frac{q(1-qA_c)^2(\frac{1}{2}-qA_c)+\theta/2}{q(1-3qA_c)+[(1+qA_c)/(1-qA_c)^2]\theta}. \tag{3.20c}$$

Integration of Eq. (3.20a) gives

$$c_{n_c}^{+}(\tau)=\frac{c_{n_c}^{+}(0)c_{n_c}^{+}(\infty)e^{\nu\tau}}{c_{n_c}^{+}(\infty)-c_{n_c}^{+}(0)(1-e^{\nu\tau})} \tag{3.21}$$

which upon substitution in Eq. (3.11) yields to first-order in ϵ

$$u(r,t)$$

$$\sim\epsilon\frac{c_{n_c}^{+}(\infty)c_{n_c}^{+}(0)\exp[-\nu(A_c-A)t]}{c_{n_c}^{+}(\infty)-c_{n_c}^{+}(0)[1-\exp[-\nu(A_c-A)t]]}\sin n_c\pi r$$

$$\begin{pmatrix}1\\ \dfrac{-q(1-qA_c)^2-\theta}{2}\end{pmatrix}+\epsilon c_{n_c}^{-}(0)e^{\sigma_{n_c}t}\begin{pmatrix}1\\ M_{n_c}^{-}\end{pmatrix}\sin n_c\pi r$$

$$+\epsilon\,\mathrm{Re}\sum_{n\neq n_c}^{\infty}c_n^{\pm}(0)e^{\sigma_n^{\pm}t}\begin{pmatrix}1\\ M_n^{\pm}\end{pmatrix}\sin n\pi r+o(\epsilon^2). \tag{3.22}$$

On Fig. 1 we have noted that the trivial solution would be asymptotically stable for values of $A>A_c$. To have such a case it must be that $c_{n_c}^{+}(\tau)=0$ for $t\to\infty$. This is obtained by setting $\nu\tau\equiv\nu(A-A_c)t<0$ at $t>0$ and $A>A_c$. Thus it follows

$$\nu=\frac{q(1-3qA_c)+(1+qA_c)\theta/(1-qA_c)^2}{2(1-\theta)}<0. \tag{3.23}$$

As $\theta<1$ we then have

$$\left\langle\hat{\Xi}_{n_c}^{+}\middle|\left[\frac{\partial L(\gamma)}{\partial\epsilon}\right]_{\epsilon=0}\hat{\Xi}_{n_c}^{+}\right\rangle$$

$$=\frac{1}{4}\left[q(1-3qA_c)+\frac{(1+qA_c)\theta}{(1-qA_c)^2}\right]<0. \tag{3.24}$$

For convenience we choose $A'(0)=1$. Thus $A-A_c=\epsilon+o(\epsilon^2)$ and to the first-order approximation,

$$x\sim c_{n_c}^{+}(\tau)(A-A_c)\sin n_c\pi r. \tag{3.25}$$

As $\nu<0$, $c_{n_c}^{+}(\infty)$ has the same sign as the quantity $q(1-qA_c)^2(\frac{1}{2}-qA_c)+\theta/2$ or equivalently, as the quantity

$$\left\langle\hat{\Xi}_{n_c}^{+}\middle|\left[\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2}N(\gamma,u)\right]_{\epsilon=0}\right\rangle$$

$$=\frac{4}{3n_c\pi}\left[q(1-qA_c)^2(\tfrac{1}{2}-qA_c)+\frac{\theta}{2}\right], \tag{3.26}$$

where $[\,]_{\epsilon=0}$ denotes a quantity evaluated at $\epsilon=0$. Thus it is clear that if $c_{n_c}^{+}(0)$ and $c_{n_c}^{+}(\infty)$ have the same sign, with $A<A_c$ the following asymptotic state is reached as $t\to\infty$

$$\begin{pmatrix}x(r)\\ y(r)\end{pmatrix}\sim\begin{pmatrix}\dfrac{A_c}{1-qA_c}\\ 1-qA_c\end{pmatrix}+\frac{16}{3n_c\pi}$$

$$\times\frac{q(1-qA_c)^2(\frac{1}{2}-qA_c)+\theta/2}{q(1-3qA_c)+(1+qA_c)\theta/(1-qA_c)^2}$$

$$\times\begin{pmatrix}1\\ \dfrac{-\theta-q(1-qA_c)^2}{2}\end{pmatrix}(A_c-A)\sin n_c\pi r+o[(A_c-A)^2], \tag{3.27a}$$

where

$$n_c^2=\left|\left|\pi^{-1}D^{-1/2}\left(\frac{A_c(1-qA_c)}{\theta}\right)^{1/4}\right|\right|. \tag{3.27b}$$

If, however, $\mathrm{sgn}c_{n_c}^{+}(0)\neq\mathrm{sgn}c_{n_c}^{+}(\infty)$, the denominator in (3.21) vanishes for a time interval of order $[\,|\nu|(A_c-A)]^{-1}$. After this time lapse the solution escapes off the ϵ region in which to a first-order is given by Eq. (3.27).

For A slightly larger than A_c, a similar analysis yields (3.27) but the branch is indeed unstable as two initially neighboring concentrations diverge as time goes on. Besides when $\mathrm{sgn}c_{n_c}^{+}(0)\neq\mathrm{sgn}c_{n_c}^{+}(\infty)$ the initial perturbation decays to the trivial fixed point which is thus asymptotically stable, whereas if $\mathrm{sgn}c_{n_c}^{+}(0)=\mathrm{sgn}c_{n_c}^{+}(\infty)$ the denominator in $c_{n_c}^{+}(\tau)$ vanishes after a time interval of order $[\,|\nu|(A-A_c)]^{-1}$. Thus the question arises: Given some initial perturbations upon the homogeneous state where does the system go when the denominator in $c_{n_c}^{+}(\tau)$ vanishes? If the possibility of the solution going to infinity is disregarded, then the system is expected to reach another steady state, say at a *finite* distance from the homogeneous fixed point as graphically depicted in Fig. 2. An argument supporting this conjecture goes as follows.

For simplicity we take the case

$$\left\langle\hat{\Xi}_{n_c}^{+}\middle|\left[\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2}N(\gamma,u)\right]_{\epsilon=0}\right\rangle<0 \tag{3.28}$$

which implies $c_{n_c}^{+}(\infty)<0$. For compactness we shall write

$$N_{\mathrm{II}}=\left[\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2}N(\gamma,u)\right]_{\epsilon=0}$$

Then for initial conditions yielding $c_{n_c}^{+}(0)$ positive on the zone of the diagram on the left of A_c the quantity $x(r,t)$ does not reach a steady state in a neighborhood of A_c of radius $\epsilon=A_c-A$. For time intervals shorter than the necessary interval for the vanishing of $c_{n_c}^{+}(\tau)$ this quantity is negative and rises with time in absolute value to eventually cross the zero value. After crossing zero the concentration becomes negative, which is not allowed here. Thus we may tentatively postulate the existence of a secondary steady state at finite distance. This can indeed be expected as the operator $L(\gamma)u+N(\gamma,u)$ is compact over the domain of interest in the Banach space F and it is also completely continuous in A on the finite segment $I_\delta=[A_c-\delta,A_c+\delta]$ for given $\delta>0$. Let Ω be the domain of the operator, and such that it encloses the postulated finite amplitude steady state for all A belonging to I_δ. Leray–Schauder's theory[11,12] gives that the topological degree is the same for all points in Ω no matter what the value of A in I_δ may be. On the other hand, at the bifurcation point there is a change in the degree. Then the following theorem can be applied[13]: Let ξ be a fixed point of a compact operator H and let L be its Fréchet differential with respect

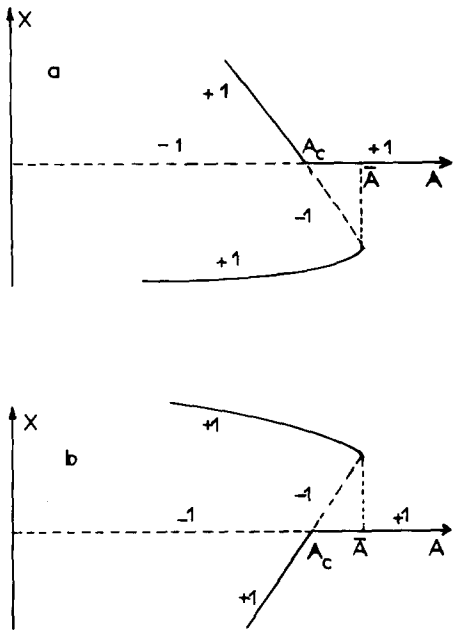

FIG. 2. Bifurcation and stability of spatial dissipative structures showing the possibility of at least one subcritical (metastable) branching. + 1 and − 1 are the corresponding Leray-Schauder degrees. (a) corresponds to the case of Eq. (3.28) and (b) to the opposite sign in (3.28).

to ξ. Assume that $\lambda = 1$ is not an eigenvalue of the problem $\phi = \lambda L\phi$. Then ξ is an isolated fixed point of H with degree $(-)^{\beta}$, where β is the sum of multiplicities of the eigenvalues of the equation given above in the unit segment [0,1]. Applying this theorem to the operator $L(\gamma)u + N(\gamma,u)$ with respect to the homogeneous fixed point $\xi = 0$, the degrees are those given in Fig. 2. Stable branches have degree + 1 whereas those unstable have − 1. Then as the topological degree in region Ω, which is the sum of all degrees for all branches and all values of A belonging to I_δ, is + 1 for $A > \bar{A}$ there must exist another branch with + 1 degree, and this justifies our assumption.

With n_c *even* to get $c_{n_c}^+(\tau)$ we ought to solve the hierarchy to the third-order equation (3.9). We have $A'(0) = 0$ and substitution of (3.11) in (3.8) and (3.9) together with

$$M_{n_c}^+ = -\tfrac{1}{2}[\theta + q(1 - qA_c)^2]$$

yield

$$\left[\frac{\partial}{\partial t} - L(\gamma_c)\right]u_2 = \begin{pmatrix} q(1-qA_c)^2(\tfrac{1}{2} - qA_c) - \theta/2 \\ \dfrac{\theta + q(1-qA_c)^2}{2} \end{pmatrix} \times c_{n_c}^{+\,2} \sin^2 n_c \pi r + (\text{e.d.t.}). \tag{3.29}$$

To find a particular solution, u^{ps}, of Eq. (3.29), we take advantage of the known eigenfunctions of $L(\gamma_c)$. We have

$$u_2^{ps} = \sum_{n \neq n_c} \alpha_n^{\pm}(\tau)\Xi_n^{\pm}(r). \tag{3.30}$$

As

$$L(\gamma_c)\Xi_n^{\pm}(r) = \sigma_n^{\pm}\Xi_n^{\pm}(r) \tag{3.31}$$

Eq. (3.29) yields

$$\sum_{n \neq n_c} \alpha_n^{\pm}(\tau)\sigma_n^{\pm}\Xi_n^{\pm}(r) = \begin{pmatrix} \theta/2 - q(1-qA_c)^2(\tfrac{1}{2} - qA_c) \\ -\tfrac{1}{2}[q(1-qA_c)^2 + \theta] \end{pmatrix} c_{n_c}^{+\,2} \sin n_c \pi r + (\text{e.d.t.}). \tag{3.32}$$

Multiplying (3.32) with $\hat{\Xi}_n^{\pm}$ with the definition (3.13) it appears that all (e.d.t.) disappear, and we obtain

$$\alpha_n^{\pm}(\tau) = \begin{cases} \rho_n^{\pm} c_{n_c}^{+\,2}(\tau), & \text{for } n \text{ odd}, \\ 0, & \text{for } n \text{ even}, \end{cases} \tag{3.33}$$

where

$$\rho_n^{\pm} \equiv \frac{8n_c^2}{n(n^2 - 4n_c^2)\pi\sigma_n^{\pm}} \times \left\{ q(1-qA_c)^2(\tfrac{1}{2} - qA_c) - \tfrac{1}{2}qA_cM_n^{\pm} - \frac{\theta}{2}\left[1 + \frac{A_cM_n^{\pm}}{(1-qA_c)^2}\right]\right\}\left[1 - \frac{A_cM_n^{\pm}}{(1-qA_c)^2}\right]^{-1}. \tag{3.34}$$

Thus the general solution of (3.29) reduces to

$$u_2 = b_{n_c}^{+}(\tau)\Xi_{n_c}^{+}(r) + c_{n_c}^{+\,2}(\tau)\Omega(r) + (\text{e.d.t.}), \tag{3.35}$$

where we have introduced the following quantity,

$$\Omega(r) \equiv \begin{pmatrix} \omega(r) \\ \zeta(r) \end{pmatrix} = \sum_{\substack{n \neq n_c \\ n \text{ odd}}} \rho_n^{\pm}\Xi_n^{\pm}(r). \tag{3.36}$$

Upon substitution of (3.11), (3.35), and (3.36) in Eqs. (3.8) and (3.9) we have

$$\left[\frac{\partial}{\partial t} - L(\gamma_c)\right]u_3 = -\tfrac{1}{2}A''(0)\frac{\partial u_1}{\partial \tau} + L_{II}(\gamma_c)u_1 + N_{III}(\gamma_c,u_1,u_2) \tag{3.37}$$

in which for compactness the following quantities have been introduced,

$$L_{II}(\gamma_c) \equiv \left[\frac{1}{2}\frac{\partial^2 L(\gamma_c)}{\partial \epsilon^2}\right]_{\epsilon = 0} = \begin{pmatrix} 1 - 2qA_c & (1-qA_c)^{-2} \\ q & -(1-qA_c)^{-2} \end{pmatrix}\frac{A''(0)}{2} \tag{3.38}$$

and

$$N_{III}(\gamma_c,u_1,u_2) = \begin{pmatrix} x_1y_2 + x_2y_1 + 2q(1-qA_c)^3x_1x_2 - q^2(1-qA_c)^4x_1^3 \\ -x_1y_2 - x_2y_1 \end{pmatrix}. \tag{3.39}$$

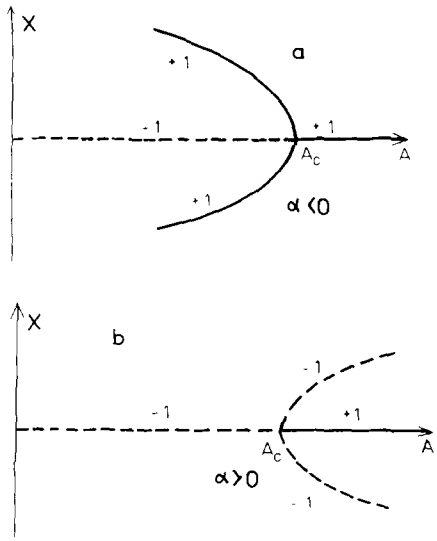

FIG. 3. Bifurcation picture corresponding to the dissipative structures given by Eq. (3.47) with the corresponding degrees. (a) and (b) denote respectively the cases $\alpha<0$ and $\alpha>0$.

Multiplication of Eq. (3.37) with $\hat{\Xi}^{+}_{n_c}(r)$ and use of Fredholm's alternative yield

$$\frac{A''(0)}{2}\frac{dc^{+}_{n_c}}{d\tau} = \frac{A''(0)}{2}\nu c^{+}_{n_c} - \alpha\nu c^{+3}_{n_c}, \tag{3.40}$$

where

$$\nu \equiv \left[1 - 2qA_c - \frac{q}{2}(1-qA_c) + \frac{1+qA_c}{(1-qA_c)^2}\frac{\theta}{2}\right]$$

$$[2(1-\theta)]^{-1}(1-\theta) \tag{3.41}$$

and

$$\alpha \equiv \left(\tfrac{3}{4}q^2(1-qA_c)^4 - \int_0^1 \left\{\left[2qA_c - \frac{A_c\theta}{(1-qA_c)^2}\right]\zeta(r) + [q(1-qA_c)^2(3-2qA_c)+\theta]\omega(r)\right\}\sin^2 n_c\pi r\,dr\right.$$

$$\left.\times\left[1 - 2qA_c - \frac{q}{2}(1-qA_c) + \frac{1+qA_c}{(1-qA_c)^2}\frac{\theta}{2}\right]^{-1}\right. . \tag{3.42}$$

In the limit $\tau\to\infty$ we find

$$c^{+}_{n_c}(\infty) = \pm[A''(0)/2\alpha]^{1/2}, \tag{3.43}$$

which upon substitution in Eq. (3.40) yields

$$\frac{dc^{+}_{n_c}}{d\tau} = \nu c^{+}_{n_c}\left[1 - \frac{c^{+2}_{n_c}(\tau)}{c^{+2}_{n_c}(\infty)}\right]. \tag{3.44}$$

Integration of (3.44) is straightforward. We have

$$c^{+}_{n_c}(\tau) = |c^{+}_{n_c}(\infty)|c^{+}_{n_c}(0)e^{\nu\tau}[c^{+2}_{n_c}(0)(e^{2\nu\tau}-1) + c^{+2}_{n_c}(\infty)]^{-1/2}. \tag{3.45}$$

It appears that $\mathrm{sgn}c^{+}_{n_c}(\tau) = \mathrm{sgn}c^{+}_{n_c}(0)$. With respect to $dc^{+}_{n_c}/d\tau$ in Eq. (3.44) two possible cases arise:

(i) If $|c^{+}_{n_c}(\infty)| < c^{+}_{n_c}(0)$ or if $-|c^{+}_{n_c}(\infty)| < c^{+}_{n_c}(0) < 0$, the slope of $c^{+}_{n_c}(\tau)$ starts being negative.

(ii) If $0 < c^{+}_{n_c}(0) < |c^{+}_{n_c}(\infty)|$ or if $-|c^{+}_{n_c}(\infty)| > c^{+}_{n_c}(0)$, the slope of $c^{+}_{n_c}(\tau)$ starts being positive.

Thus if $c^{+}_{n_c}(0)$ is positive, then $c^{+}_{n_c}(\tau)$ goes monotonically to $|c^{+}_{n_c}(\infty)|$ whereas if $c^{+}_{n_c}(0)$ is negative, then $c^{+}_{n_c}(\tau)$ goes monotonically to $-|c^{+}_{n_c}(\infty)|$. This shows that the dissipative structure found at $t\to\infty$ depends only on the sign of the initial conditions.

In this case the asymptotic expansion of the solution is

$$\begin{pmatrix}x(r,t,\epsilon)\\ y(r,t,\epsilon)\end{pmatrix} = \begin{pmatrix}A_c/(1-qA_c)\\ (1-qA_c)\end{pmatrix} + \epsilon\left(\begin{array}{c}1\\ \dfrac{-\theta - q(1-qA_c)^2}{2}\end{array}\right)\times\frac{c^{+}_{n_c}(0)|c^{+}_{n_c}(\infty)|\exp[\nu(A-A_c)t]}{[c^{+2}_{n_c}(0)\{\exp[2\nu(A-A_c)t]-1\}+c^{+2}_{n_c}(\infty)]^{1/2}}$$

$$\times\sin n_c\pi r + \epsilon c^{-}_{n_c}(0)\begin{pmatrix}1\\ M^{-}_{n_c}\end{pmatrix}e^{\sigma_{n_c}t}\sin n_c\pi r + \epsilon\sum_{n\neq n_c}^{\infty} c^{\pm}_{n}(0)e^{\sigma_n t}\begin{pmatrix}1\\ M^{\pm}_{n}\end{pmatrix}\sin n\pi r + o(\epsilon^2), \tag{3.46}$$

which in the limit $t\to\infty$ yields depending on the i.c. either one or the other of the following two steady inhomogeneous solutions,

$$\begin{pmatrix}x(r)\\ y(r)\end{pmatrix} \sim \begin{pmatrix}A_c/(1-qA_c)\\ 1-qA_c\end{pmatrix} \pm \left(\frac{A-A_c}{\alpha}\right)^{1/2}\times\left(\frac{-\theta^{1}+q(1-qA_c)^2}{2}\right)\sin n_c\pi r + o(|A-A_c|). \tag{3.47}$$

As for $A>A_c$, the trivial solution is asymptotically stable, then we also have here

$$\nu<0 \tag{3.48}$$

as it was found in the case of n_c odd.

Notice that if the ϵ expansion is to be valid we must have $c^{+2}_{n_c}(\infty)$ positive. Thus:

(i) If $\alpha<0$, (3.43) implies $A''(0)<0$ and $A<A_c$, the dissipative structures are asymptotically stable.

(ii) If $\alpha>0$, $A''(0)>0$, and $A>A_c$, the inhomogeneous steady states are unstable. In this case $\exp[\nu(A-A_c)t]\to 0$, and $c^{+}_{n_c}(\tau)\to 0$, and the inhomogeneous solutions decay to the homogeneous fixed point in a time interval of order $[|\nu|(A-A_c)]^{-1}$. The corresponding bifurcation picture is given in Fig. 3.

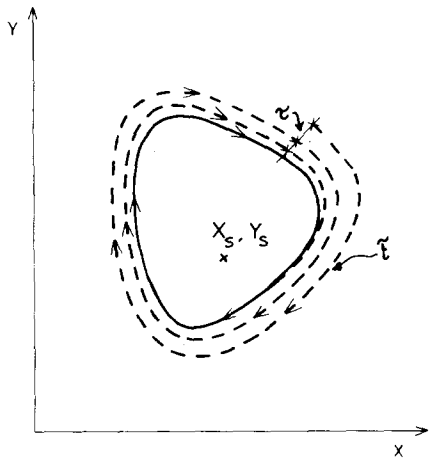

FIG. 4. The limit cycle around (1.4)—the steady state X_s, Y_s—and the two time-scales introduced in Eq. (4.4). $\tilde{t}$ proceeds with the actual trajectory of the initial condition until the system reaches the limit cycle, whereas τ, the slower scale transversally crosses the trajectory from the initial condition down to the limit cycle.

4. LIMIT CYCLE BEHAVIOR

In this section we shall construct the limit cycle that bifurcates from the unstable trivial fixed point in the neighborhood of the neutral stability curve (q,A) next to region II (Fig. 1). Again our analysis uses the two time-scales[10] whose physical interpretation is made transparent in Fig. 4.

The neutral stability curve to be considered in Eq. (2.12),

$$qA_c(1-qA_c) - \frac{A_c}{1-qA_c} = \pi^2 D(1+\theta), \tag{4.1}$$

and the critical eigenvalue is $n_c = 1$. We have

$$\sigma_1^{\pm} A_c = \pm i\omega$$

$$\equiv \pm i(A_c - [qA_c(1-qA_c) - \pi^2\theta D\,]^2)^{1/2}. \tag{4.2}$$

From Eqs. (2.10) we get the constants $M_1^{\pm}$. They are given by

$$\pm i\omega - qA_c(1-qA_c) + \pi^2 D\theta - \frac{A_c}{1-qA_c} M_1^{\pm} = 0. \tag{4.3}$$

For the concentrations the following solution is assumed,

$$u_1(r,t,\tau) = \mathrm{Re}\{c_1^+(\tau)e^{i\omega t}\Xi_1^+(r) + c_1^-(\tau)e^{-i\omega t}\Xi_1^-(r)\}$$

$$+ \text{(e.d.t.)}, \tag{4.4}$$

where t and τ are the two time scales (Fig. 4).

Defining

$$c_1(\tau) \equiv \tfrac{1}{2}[c_1^+(\tau) + c_1^{-*}(\tau)]$$

and using $M_1^- = M_1^{+*}$ together with $e^{i\omega t}\Xi_1^-(r) = [e^{i\omega t}\Xi_1^+(r)]^*$ and (4.4) we get

$$u_1(r,t,\tau) = c_1(\tau)e^{i\omega t}\Xi_1^+(r) + \text{c.c.} + \text{(e.d.t.)}, \tag{4.5}$$

where c.c. accounts for complex conjugation of the preceding term. The following i.c. for $c_1(\tau)$ comes from Eq. (3.12),

$$c_1(0) = 2\int_0^1 \left\{h_\epsilon(r,0) - \frac{A_c M_1^+}{(1-qA_c)^2} g_\epsilon(r,0)\right\} \sin\pi r\, dr$$

$$\times [1 - A_c(1-qA_c)^{-2} M_1^{+2}]^{-1}. \tag{4.6}$$

Substituting (4.5) in Eq. (3.8) we have:

$$\left[\frac{\partial}{\partial t} - qA_c(1-qA_c) - \theta D\frac{\partial^2}{\partial r^2}\right]x_2 - \frac{A_c}{1-qA_c}y_2$$

$$= -A'(0)\sin\pi r(c_1' e^{i\omega t} + \text{c.c.}) + q(1-2qA_c)A'(0)$$

$$\times \sin\pi r(c_1 e^{i\omega t} + \text{c.c}) + \frac{A'(0)}{(1-qA_c)^2}\sin\pi r(c_1 M_1^+ e^{i\omega t}$$

$$+ \text{c.c}) + \sin^2\pi r[(M_1^+ + M_1^{+*})|c_1|^2 + c_1^2 e^{i2\omega t}$$

$$+ \text{c.c.}] + q(1-qA_c)^3\sin^2\pi r(2|c_1|^2 + c_1^2 e^{i2\omega t}$$

$$+ \text{c.c.}) + \text{(e.d.t.)}, \tag{4.7a}$$

$$\left[\frac{\partial}{\partial t} + \frac{A_c}{1-qA_c} - D\frac{\partial^2}{\partial r^2}\right]y_2 + (1-qA_c)x_2$$

$$= -A'(0)\sin\pi r(c_1' M_1^+ e^{i\omega t} + \text{c.c.}) + qA'(0)\sin\pi r(c_1 e^{i\omega t}$$

$$+ \text{c.c.}) - \frac{A'(0)}{(1-qA_c)^2}\sin\pi r(c_1 M_1^+ e^{i\omega t} + \text{c.c.})$$

$$- \sin^2\pi r[(M_1^+ + M_1^{+*})|c_1|^2 + c_1^2 M_1^+ e^{i2\omega t}$$

$$+ \text{c.c.}] + \text{(e.d.t.)}. \tag{4.7b}$$

Defining for later convenience the following average

$$\langle\langle[\hat{\Xi}_1^+]\,|f\rangle\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle\hat{\Xi}_1^+|f\rangle e^{i\omega t}\,dt \tag{4.8}$$

and using Fredholm's alternative with (4.7) we get $A'(0) = 0$ if c_1 is not to be vanishing. Substitution of this result in Eq. (4.7) and integration of the system yields

$$u_2(r,t,\tau) = b_1(\tau)e^{i\omega t}\Xi_1^+(r) + \text{c.c.} + c_1^2(\tau)e^{i2\omega t}\Omega(r) + \text{c.c.}$$

$$+ |c_1(\tau)|^2[\Omega(r) + \text{c.c.}]_{\omega=0} + \text{(e.d.t.)} \tag{4.9}$$

in which

$$\Omega(r) = \begin{pmatrix}\omega(r)\\ \zeta(r)\end{pmatrix}$$

$$= \sum_{n=3}^{\infty}\frac{\rho_n^{\pm}}{iz\omega - \sigma_n^{\pm}}$$

$$\times\left\{M_1^+\left[1 + \frac{A_c M_n^{\pm}}{(1-qA_c)^2}\right] + q(1-qA_c)^3\right\}\Xi_n^{\pm}(r), \tag{4.10}$$

with

$$\rho_n^{\pm} = \frac{-8}{n(n^2-4)}\left\{1 - \frac{A_c M_n^{\pm 2}}{1-qA_c}\right\}^{-1}. \tag{4.11}$$

a bracket $[\,]_{\omega=0}$ will denote the use of $-\rho_n^{\pm}/\sigma_n^{\pm}$ instead of $\rho_n^{\pm}/(i2\omega-\sigma_n^{\pm})$.

To obtain $c_1(\tau)$ we introduce Eq. (4.9) into the third order Eq. (3.9). Multiplication of the result with $\hat{\Xi}_1^{+}(r)$ and use of (4.8) yields the following differential equation,

$$\frac{A''(0)}{2}\frac{dc_1^*}{d\tau} = \frac{A''(0)}{2}\nu c_1^* + \lambda\,|c_1|^2 c_1^*, \tag{4.12}$$

where

$$\nu = \left\{q(1-2qA_c) + \frac{M_1^{+*}}{1-qA_c} + \frac{A_c M_1^{+*}}{(1-qA_c)^4}\right\} \times\left\{1 - \frac{A_c M^{+*}}{(1-qA_c)^2}\right\}^{-1}, \tag{4.13}$$

$$\lambda = 2\int_0^1 \sin^2\pi r\, dr\{2q(1-qA_c)\omega^*(r) + [\omega(r)+\text{c.c.}]_{\omega=0} + \left[1 + M_1^{+*} \times \frac{A_c}{(1-qA_c)^2}\right](\zeta^*(r) + [\zeta(r)+\text{c.c.}]_{\omega=0} + M_1^{+}\omega^*(r) + M_1^{+*}[\omega(r)+\text{c.c.}]_{\omega=0})\Big] \times[1 - A_c M_1^{+*}(1-qA_c)^{-2}]^{-1}. \tag{4.14}$$

From (4.12) we take $c_1(\tau) = c(\tau)e^{-i\alpha(\tau)}$, with $c(\tau)$ and $\alpha(\tau)$ being some functions of τ yet to be specified. Then separation of real and imaginary parts in (4.12) leads to the following two equations:

$$\frac{A''(0)}{2}\frac{dc}{d\tau} = \frac{A''(0)}{2}c\,\text{Re}\nu + c^3\,\text{Re}\lambda, \tag{4.15a}$$

$$\frac{A''(0)}{2}\frac{d\alpha}{d\tau} = \frac{A''(0)}{2}\text{Im}\nu + c^2\,\text{Im}\lambda. \tag{4.15b}$$

Thus from (4.15a) at $\tau\to\infty$ we get

$$c(\infty) = \left[-\frac{A''(0)}{2}\frac{\text{Re}\nu}{\text{Re}\lambda}\right]^{1/2}, \tag{4.16}$$

which upon substitution in Equation (4.15a) yields

$$\frac{dc}{d\tau} = \text{Re}\nu\left[1 - \frac{c^2}{c(\infty)^2}\right]c. \tag{4.17}$$

From Sec. 2 the solution to (4.17) is known. We have

$$c(\tau) = \frac{c(0)c(\infty)e^{\tau\,\text{Re}\nu}}{\{c(\infty)^2 + c(0)^2[e^{2\tau\,\text{Re}\nu}-1]\}^{1/2}}. \tag{4.18}$$

The unknown phase follows from (4.15b) and (4.18),

$$\alpha(\tau) = \alpha(0) + \tau\,\text{Im}\nu + \frac{2\,\text{Im}\lambda}{A''(0)}\int_0^{\tau} c^2(s)\,ds, \tag{4.19}$$

which for large values of time yields the following behavior,

$$\alpha(\tau)\sim\tau\left\{\text{Im}\nu - \frac{\text{Im}\lambda\,\text{Re}\nu}{\text{Re}\lambda}\right\}. \tag{4.20}$$

Thus to the first order in ϵ the following result is obtained,

$$\begin{aligned} u(r,t,\epsilon) &\sim \epsilon 2c(0)c(\infty)\exp[-(A_c-A)\,\text{Re}\nu t\,] \\ &\quad\times[c(\infty)^2 + c(0)^2(\exp[-2(A_c-A)\,\text{Re}\nu t\,]-1)]^{-1/2} \\ &\quad\times\left(\frac{1}{\frac{2\pi^2\theta D}{A_c}(1-qA_c) - 2q(1-qA_c)^2}\right) \\ &\quad\times\cos[\omega t - \alpha(A-A_c)t\,] + \epsilon\,\text{Re}\sum_{n=2}^{\infty} c_{n_c}^{\pm}(0) \\ &\quad\times e^{\sigma_{n_c}^{\pm}t}\Xi_{n_c}^{\pm}(r) + o(\epsilon^2), \end{aligned} \tag{4.21}$$

which for large values of time, i.e., asymptotically leads to the time periodic solution

$$\begin{aligned} \begin{pmatrix} X(r,t) \\ Y(r,t) \end{pmatrix} &\sim \begin{pmatrix} \frac{A_c}{1-qA_c} \\ 1-qA_c \end{pmatrix} + 2\left[\frac{\text{Re}\nu}{\text{Re}\lambda}(A_c-A)\right]^{1/2}\sin\pi r \\ &\quad\times\left(\frac{1}{2\left[\pi^2\theta D\frac{1-qA_c}{A_c} - q(1-qA_c)^2\right]}\right) \\ &\quad\times\cos\left[\omega + (A_c-A)\left(\text{Im}\nu - \frac{\text{Im}\lambda\,\text{Re}\nu}{\text{Re}\lambda}\right)\right]t \\ &\quad + o(|A-A_c|). \end{aligned} \tag{4.22}$$

For $A > A_c$ this trivial solution (1.4) is asymptotically stable and will occur for $\text{Re}\nu < 0$. Thus as $c(\infty)$ must be real there is the following alternative: Either

(i) $\text{Re}\lambda < 0$. We have $A < A_c$ and the solution (4.22) is the expected limit cycle. Or

(ii) $\text{Re}\lambda > 0$. Then $A > A_c$ and the (orbit) time periodic solution (4.22) is unstable.

It is clear that if the diffusion parameters are large enough the limit cycle behavior may not be seen. From Eq. (4.2) we note that if D_X is so large as to allow only vanishing values of ω, the eigenvalues $\sigma_1^{\pm}(A = A_c)$ and (2.14) cross each other and for $A < A_c$ we have no bifurcation to a limit cycle but instead we have branching of at least two nonuniform steady states. For with $\omega = 0$ there is $M_1^{+} = M_1^{-}$ and $2c_1(\tau) = c_{n_c}^{+}(\tau)$ and the solution of this Sec. 4 coincides with

(3.11) for n_c odd (and equal to one). Then as already discussed, in the preceding Sec. 3, of these two expected steady solutions one grows with amplitude continuously varying on $|A_c - A|$, and the other ought to be located at a finite distance from the critical point (subcritical or metastable branching). In conclusion, Hopf bifurcation to a limit cycle is expected along the neutral stability curve (4.1) provided

$$1 < q < Q \equiv [1 + (1 + 4A_c^{1/2} - 4\pi^2\theta D)^{1/2}]/2A_c.$$

5. CONCLUDING REMARKS

Perhaps the most interesting feature of the method used, with the two time-scales defined, is that it not only provides the construction of the nonlinear bifurcated branches but that it also yields information on their stability, and permits us to see in detail the evolution of initial disturbances upon the trivial steady state of the system. According to Fig. 1 this fixed point is unstable for $A \leqslant A_c$, and at least locally asymptotically stable otherwise. Thus the solution $u_1(r,t,\tau)$ must go to zero as time goes to infinity, namely $u_1 \to 0$ as $\tau = (A - A_c)t \to \infty$, provided $A > A_c$. Then we have[2]:

(i) with n_c odd, Eq. (3.21) belonging to the spatial dissipative structure (sds) demands that ν be negative,

(ii) with n_c even, Eq. (3.45) again for a sds demands that ν be negative,

(iii) Eq. (4.18) for the temporal dissipative structure (the limit cycle) demands that $\mathrm{Re}\nu$ be negative.

The time evolution of initial disturbances is given for the three cases by the respective Eqs. (3.22), (3.46), and (4.21), from which the following conclusions are inferred:

(i) with n_c odd, the sds is semistable for $A < A_c$ and unstable otherwise, and two possibilities exist (see Fig. 2) according to the sign of $\langle \Xi_{n_c}^{+} \mid N_{11} \rangle$. The corresponding relaxation time is of order $[\nu(A - A_c)]^{-1}$,

(ii) with n_c even, the sds is asymptotically stable for $\alpha < 0$ and unstable otherwise,

(iii) the limit cycle is orbitally asymptotically stable for $\mathrm{Re}\lambda < 0$ and unstable otherwise. Its relaxation time is of order $[\mathrm{Re}\nu(A - A_c)]^{-1}$. The method has also permitted us to locate a subcritical branch at a finite distance as shown in Fig. 2.

ACKNOWLEDGMENTS

The results reported in this note were obtained in the course of research sponsored by the Instituto de Estudios Nucleares (Spain). L.L. Bonilla is also grateful to the Ministerio de Educación y Ciencia (Spain) for a predoctoral fellowship. Both authors acknowledge fruitful discussions with J.L. Ibáñez.

[1] J.L. Ibáñez and M.G. Velarde, J. Math. Phys. **19**, 151 (1978).
[2] J.L. Ibáñez and M.G. Velarde, Phys. Rev. A **18**, 750 (1978).
[3] I. Balslev and H. Degan, J. Theor. Biol. **49**, 173 (1975).
[4] R. Aris, *The Mathematical Theory of Diffusion and Reaction in Permeable Catalysts* (Clarendon, Oxford, 1976), Vol. I, Sec. 4. 7.3, Vol. II, Secs. 7.9 and 8.8.
[5] J.L. Ibáñez, V, Fairén, and M.G. Velarde, Phys. Lett. A **58**, 364 (1976).
[6] P. Glansdorff and I. Prigogine, *Thermodynamics of Structure, Stability and Fluctuations* (Wiley-Interscience, New York, 1971).
[7] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Interscience-Wiley, New York, 1977).
[8] H. Haken, *Synergetics: An Introduction* (Springer-Verlag, Berlin, Heidelberg, New York, 1977).
[9] J.L. Ibáñez, Ph.D. dissertation (Universidad Autónoma de Madrid, 1977) unpublished.
[10] S. Kogelman and H.B. Keller, SIAM J. Appl. Math. **20**, 619 (1971).
[11] D.H. Sattinger, *Topics in Stability and Bifurcation Theory* (Springer-Verlag, Berlin, Heidelberg, New York, 1973).
[12] P.H. Rabinowitz, Lectures on nonlinear problems, University of Paris-VI. (1973), unpublished report No. 75010 (1975).
[13] G.R. Gavalas, *Non-Linear Differential Equations of Chemically Reacting Systems* (Springer-Verlag, Berlin, Heidelberg, New York, 1968).
[14] J.E. Marsden and M. MacCracken, *The Hopf Bifurcation and Its Applications* (Springer-Verlag, Berlin, Heidelberg, New York, 1976).

ERRATA

Erratum: Higher Euler operators and some of their applications [J. Math. Phys. 20, 522 (1979)]

S. J. Aldersley
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Reference 19 was omitted from the list of footnotes. It is: S.J. Aldersley, "Comments on certain divergence-free tensor densities in a three-space," to appear in J. Math. Phys.

Erratum: Asymptotic forms of radial wavefunctions and Jost functions for cutoff potentials [J. Math. Phys. 20, 1210 (1979)]

W. J. Romo
Physics Department, Carleton University, Ottawa, Ontario K1S 5B6

P.1214, Eq. (2.28e): replace $\mathscr{M}^{j}_{\alpha+\alpha}(k)$ by $\mathscr{M}^{j}_{\alpha,-\alpha}(k)$; last line of Eq. (2.29): replace the exponent of k by $(-\mu_q-\mu_r-4)$.

P.1215, Eq. (2.36): replace the factor $(-i\alpha k)^{-l-1}$ on the right-hand side of the equation by $\Sigma_{\alpha=-1}^{+1}(i\alpha k)^{-l-1}$.

P.1218, Eq. (3.10): replace the exponent of z on the last line of the equation by $\alpha_{N(j)}-\alpha_j$

P.1223, nine lines below Eq. (B3): replace $\alpha=\pm 1$ by $\alpha=+1$; relabel Eqs. (B12a) and (B12b) as Eqs. (B11c) and (B11d), respectively.

P.1216, Eq. (2.44b): replace j^l by i^l.

 0022-2488/79/122704-01$00.00